Abstract Title Page

Not included in page count.

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

Authors and Affiliations:

Dan Goldhaber, Cyrus Grout, and Nick Huntington-Klein
dgoldhab@uw.edu, cyrusgrout@gmail.com, nickck@uw.edu
Center for Education Data and Research
University of Washington

Abstract Body

Limit 4 pages single-spaced.

Background / Context:

It is well documented that teachers can have profound effects on student outcomes. Empirical estimates find that a one standard deviation increase in teacher quality raises student test achievement by 10 to 25 percent of a standard deviation.[*] More recent evidence shows that the effectiveness of teachers can affect long-term student outcomes, such as college-going behavior and labor earnings (Chamberlain, 2013; Chetty, Friedman, & Rockoff, forthcoming; Jackson, 2012; Koedel, 2008). 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, including efforts to increase teacher effectiveness through alternative certification, new processes of evaluation and feedback, professional development, the provision of performance incentives, and, more recently, a focus on pre-service training providers.[†]

There exists far less research or policy focus on the choices school systems make in the teacher hiring process, and what tools or policies may aid the process. This is surprising for several reasons. First, there is a large and growing body of economic research pointing to the importance of employee selection, both in general (Shaw & Lazear, 2007) and in reference to screening and selection tools specifically (McDaniel et al. 1988; Gandy et al., 1994; Bliesener, 1996). Second, many school districts often have a fair amount of choice over who they might hire, but once hired (and particularly if tenured), it can be quite costly to remove a teacher seen as ineffective (National Council on Teacher Quality, 2014; Treu, 2014). Third, the credentials that are generally used to determine employment eligibility tend to be only weakly correlated with teacher effectiveness.[‡] Districts may be making a large long-term financial commitment with large effects on achievement when a teacher is hired. The recruitment and selection process should work well.

Purpose / Objective / Research Question / Focus of Study:

We analyze the relationship between two teacher selection rubrics that are used during the teacher hiring process in Spokane Public Schools (SPS). We are interested in whether these screening scores are capable of predicting strong teacher outcomes: value-added measures of effectiveness, teacher absence behavior, and the likelihood of attrition. All three of these measures are arguably quite important. Value-added measures of teacher performance have been found to be predictive of students' future test achievement (Kane & Staiger, 2008; Goldhaber & Hansen, 2013) and long-term outcomes (Chetty et al., forthcoming). Evidence suggests that teacher absences are related to their value-added (Clotfelter, Ladd, & Vigdor, 2009; Herrmann and Rockoff, 2011; Miller, Murnane, & Willet, 2008) and may also have broader impacts on students and schools (Clotfelter, Ladd, & Vigdor, 2009). Teacher attrition has important

[*] See, for instance, Aaronson et al. (2007), Goldhaber and Hansen (2013), and Hanushek and Rivkin (2010) for estimates of the effect size associated with changes in teacher quality.

[†] See, for example: Glazerman et al. (2006) and Goldhaber and Brewer (2000) 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 (2013) on teacher preparation; and Garet et al. (2001; 2011) on professional development.

[‡] See, for instance, Aaronson et al. (2007), Goldhaber and Hansen (2013) , Harris and Sass (2011), and Rivkin et al. (2005).

implications for both district administrative costs and student achievement.[§] Strong predictive validity would lend support to the use of screening tools for teacher selection. In particular, since the screening scores we analyze use qualitative data scored by a standardized coding system, strong predictive validity would lend support to the use of data that is human-interpreted in a standardized manner.

Our study is unique in several aspects. Unlike previous studies of hiring, we observe employment outcomes for both applicants who are hired into SPS *and* applicants who are not hired into SPS, some of whom are observed working elsewhere in the state. Without these observations, analysis cannot speak to the ability of the screening process to accurately separate high-end teachers from low-end teachers. Additionally, we are able to obtain credible causal estimates of the validity of the screening process by correcting for selection into the sample.

Setting:

We analyze the relationship between two teacher selection rubrics that are used during the teacher hiring process in Spokane Public Schools (SPS). Spokane is the second-largest district in Washington state, and the largest in eastern Washington.

Population / Participants / Subjects:

We observe the hiring process from 2009-2012, during which 2,669 individuals applied to 521 classroom teaching positions for which at least one teacher was screened.

Intervention / Program / Practice:

Spokane uses a four-stage hiring process. First, job applications are submitted through an online applicant management system. Those with complete information are given a screening score on a 21-point rubric with three components[**] by the district's human resources department, which is not specific to a particular position. Second, principals detail applicant requirements (i.e. "send all applications with screening scores above 17"). Qualifying applicant files sent to the school are given a screening score on a 60-point rubric with ten components,[††] scored with the particular position in mind. Both rubrics are based on information from applicants' letters of recommendation. Third, applicants with the best 60-point screening scores are interviewed, and the final hiring decision is made based on the interview.

Research Design:

We perform a series of linear and logit regression analyses, using teacher performance (student performance on reading or math exams, for teachers in the appropriate disciplines, teacher absences, and attrition from the school, district, or profession) as the dependent variable, and controls for student characteristics (in the student achievement model), teacher characteristics (in the absence and attrition models), and performance on the screening score measures. For example, the student achievement model is specified as:

$$Y_{ijsgt} = \alpha_0 + \alpha_1 Y_{i(g-1)(t-1)} + \alpha_2 X_{igt} + \alpha_3 SCREEN_{j(t-1)} + \alpha_g + \alpha_t + \varepsilon_{ijst}^{\alpha}$$

[§] Recent evidence, for instance, pegs the cost of teacher recruitment and selection at over $1000 per teacher (Milanowski & Odden, 2007). See (Ronfeldt, Loeb, & Wyckoff, 2013) on how turnover in schools affects student achievement.

[**] Experience, Skills, and Letters of Recommendation.

[††] Certificate & Education, Training, Experience, Classroom Management, Flexibility, Instructional Skills, Interpersonal Skills, Cultural Competency, Preferred Qualifications, and Letters of Recommendation.

Where Y_{ijsgt} is the state 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 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/reduced price lunch, English language learner status, gifted status, special education status, learning disability status, migrant status, and homeless status); and $SCREEN_j$ is the screening score for teacher ***j***. In all models, standard errors are clustered at the teacher/year level.

$SCREEN_j$ enters the model in a number of different ways. In specification 1, $SCREEN_j$ is the teacher's standardized total score on the 21-point pre-screen. In specification 2, $SCREEN_j$ is one of the standardized components of the 21-point pre-screen scores. In specification 3, $SCREEN_j$ is the teacher's standardized rater total on the 60-point screen. Specification 4 is analogous to specification 2, but includes each component of the adjusted 60-point screening score in $SCREEN_j$.[‡‡] In specification 5, $SCREEN_j$ includes the teacher's total scores on both the rater total for the 21-point and 60-point screening scores.

In addition to the standard regression analyses, we perform a Heckman (1979) correction for selection into the sample, using the selection equation

$$Hired^* = \alpha_{H0} + \alpha_{H1}Z_t + \alpha_{H2}SCREEN_j + \varepsilon_t^H, Hired = I(Hired^* \geq 0)$$

Where $Hired^*$ is the applicant's propensity to be hired ($Hired = 1$), estimated with probit. $SCREEN_j$ is the applicant's 21-point screening score. Z_t are variables which affect the hiring process but are otherwise unrelated to teacher quality. These are the quality of competition faced by the applicant (21-point screening scores of other applicants for the job), and an indicator for an erroneously high screening score due to an arithmetic error.

Data Collection and Analysis:

We link data at the student, teacher, and school level. Student background, test score, and teacher assignment data come from the statewide Core Student Record System (CSRS). The state's annual assessment of student learning is administered annually to students in grades 3–8. Teacher and applicant data come from multiple sources. SPS provides information on all applicants, including records of which jobs each applicant applied to, applicant characteristics, screening instrument scores, and information on teacher absences for those who are hired into Spokane.

Spokane applicants are linked to statewide teacher data sets. Professional Education and Standards Board (PESB) data include licensure test scores and endorsement areas. Teacher absence data for teachers who do not work in Spokane comes from the Washington School Information Processing Cooperative (WSPIC). We also link applicants to the S-275, a personnel report of all certificated employees of school districts in Washington State, with demographic information, experience, and contract information.

The average characteristics of applicants as they move through the hiring process are in in **Table 1**. Note that since some teachers apply in multiple years, the total number of observations is greater than 2,669. Teacher characteristics improve as weaker teachers are screened out during the hiring process.

[‡‡] We estimate specifications 2 and 4 separately for each sub-component, to avoid issues of collinearity between components.

Findings / Results:
The relationship between the screening scores and teacher performance are detailed in **Tables 2-4**, respectively. Coefficients are the effect of a one standard deviation change in an applicant's score or the listed subcomponent on standardized test scores, days of absence, or the log odds of attriting, respectively..

Applicant scores on the 21-point rubric have a positive relationship with student achievement in both math and reading (specification 1, the top row in **Table 2**), but the relationship is only significant for math. The relationship between the 60-point rubric and student achievement (specification 3) is more consistently significantly positive for both subjects, and arguably educationally significant. A 60-point screening score increase of one standard deviation associated with math achievement 8-9% of a standard deviation higher and reading achievement 5-6% of a standard deviation higher.

In **Table 3**, the screening scores do not have a significant effect on teacher absences, and few of the subcomponents have significant effects. The magnitude of these non-significant effects appears to be due to a correlation between absences and experience (e.g., Clotfelter et al., 2011). Controlling for experience causes these coefficients to shrink.

In **Table 4**, screening scores are associated with significantly lower levels of teacher attrition. 60-point screening scores are particularly predictive. Notably, Interpersonal Skills, insignificant for other outcomes, significantly predicts attrition.

Sample selection-corrected estimates for Specification 5 for all three performance outcomes are in **Table 5**.[§§] In no case is the difference between the corrected and uncorrected estimates statistically or meaningfully different, suggesting that sample selection bias is not a major concern in this particular case.

Conclusions:
We find that screening scores are valid predictors of student achievement and attrition, although not absences, in Spokane. Teacher hiring processed in general could be improved by similar screening processes, although it does not necessarily follow that these results can be generalized beyond Spokane, a district that is often seen as a desirable place to work in Eastern Washington. Spokane may not face the same hiring problems as other districts. However, these results are consistent with the wider literature on screening at the hiring stage in other industries. The screening scores used by SPS largely represent the guided human interpretation of subjective letters of recommendation, and provide information that predicts which teachers perform well. Given alternate literature on screening in hiring (e.g., Ebmeier & Ng, 2006; Oyer & Schaefer, 2011), we may expect that this guided and codified interpretation of subjective data improves on the ad-hoc hiring processes typically seen in public schools.

[§§] Note that the "uncorrected" estimates in **Table 5** differ from those in **Tables 2-4** since here the sample is necessarily limited to those hired into Spokane, and not those who end up working elsewhere.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago Public High Schools. *Journal of Labor Economics*, *25*(1), 95–135.

Bliesener, T. (1996). Methodological Moderators in Validating Biographical Data in Personnel Selection. *Journal of Occupational and Organizational Psychology*, *69*(1), 107–120.

Chamberlain, G. E. (2013). Predictive effects of teachers and schools on test scores, college attendance, and earnings. *Proceedings of the National Academy of ...*.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (9999). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*.

Clotfelter, C. T., Ladd, H., & Vigdor, J. L. (2009). Are Teacher Absences Worth Worrying About in the United States? *Education Finance and Policy*, *4*(2), 115–149.

Dee, T., & Wyckoff, J. (2013). Incentives, Selection, and Teacher Performance: Evidence from IMPACT. *NBER Working Paper 19529*.

Ebmeier, H., & Ng, J. (2006). Development and Field Test of an Employment Selection Instrument for Teachers in Urban School Districts. *Journal of Personnel Evaluation in Education2*, *18*(3), 201–218.

Gandy, J. A., Dye, D. A., & MacLane, C. N. (1994). Federal Government Selection: The Individual Achievement Record. In G. S. Stokes, M. D. Mumford, & W. A. Owens (Eds.), *Biodata Handbook: Theory, Research, and Use of Biographical Information in Selection and Performance Prediction* (xxii., pp. 275–309). Palo Alto, CA: CPP Books.

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What Makes Professional Development Effective? Results from a National Sample of Teachers. *American Education Research Journal*, *38*(4), 915–945.

Garet, M. S., Wayne, A. J., Stancavage, F., Taylor, J., Eaton, M., Walters, K., … Warner, E. (2011). *Middle School Mathematics Professional Development Impact Study: Findings After the Second Year of Implementation* (p. 205). Institute for Education Sciences, Washington, DC.

Glazerman, S., Mayer, D. P., & Decker, P. T. (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, D. (2013). *What Do Value-Added Measures of Teacher Preparation Programs Tell Us? Carnegie Knowledge Network.*

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.

Goldhaber, D., & Hansen, M. (2013). Is it Just a Bad Class? Assessing the Long-term Stability of Estimated Teacher Performance. *Economica*, *80*(319), 589–612.

Hanushek, E. A., & Rivkin, S. G. (2010). Generalizations About Using Value-added Measures of Teacher Quality. *American Economic Review*, *100*(2), 267–271.

Harris, D. N., & Sass, T. R. (2011). Teacher Training, Teacher Quality and Student Achievement. *Journal of Public Economics*, *95*(7-8), 798–812. doi:10.1016/j.jpubeco.2010.11.009

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, *47*(1), 153–161.

Herrmann, M. A., & Rockoff, J. E. (2012). Worker Absence and Productivity: Evidence from Teaching. *Journal of Labor Economics*, *30*(4), 749–782.

Hill, H. C., & Grossman, P. (2013). Challenges and Opportunities Posed by New Teacher Evaluation Systems. *Harvard Educational Review*, *83*(2), 371–385.

Jackson, C. K. (2012). Non-cognitive ability, test scores, and teacher quality: Evidence from 9th grade teachers in North Carolina.

Kane, T. J., & Staiger, D. O. (2008). *Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation* (No. Working Paper #14607).

Koedel, C. (2008). Teacher quality and dropout outcomes in a large, urban school district. *Journal of Urban Economics*.

McDaniel, M. A., Schmidt, F. L., & Hunter, J. E. (1988). A Meta-analysis of the Validity of Methods for Rating Training and Experience in Personnel Selection. *Personnel Psychology*, *41*(2), 283–309.

Milanowski, A. T., & Odden, A. R. (2007). *A New Approach to the Cost of Teacher Turnover* (No. WP-13). *The School Finance Redesign Project (SFRP) Working Paper Series*. University of Washington, Seattle, WA: Center on Revinventing Public Education.

Miller, R. T., Murnane, R. J., & Willet, 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.

National Council on Teacher Quality. (2014). *2013 State Teacher Policy Yearbook: National Summary*.

Neal, D. (2011). *The Design of Performance Pay in Education* (No. 16710). Cambridge, MA.

Oyer, P., & Schaefer, S. (2011). Personnel Economics: Hiring and Incentives. In D. Car & O. Ashenfelter (Eds.), *Handbook of Labor Economics* (Volume 4., pp. 1769–1823). Elsevier B.V.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, Schools, and Academic Achievement. *Econometrica*, *73*(2), 417–458.

Ronfeldt, M., Loeb, S., & Wyckoff, J. H. (2013). How Teacher Turnover Harms Student Achievement. *American Educational Research Journal*, *50*(1), 4–36.

Shaw, K., & Lazear, E. (2007). Personnel Economics: The Economist's View of Human Resources. *Journal of Economic Perspectives*, *21*(4), 91–114.

Springer, M. G., Ballou, D., Hamilton, L. S., Le, V.-N., Lockwood, J. R., McCaffrey, D. F., Stecher, B. M. (2010). Teacher Pay For Performance: Experimental Evidence from the Project on Incentives in Teaching. Santa Monica, CA: RAND Corporation.

Treu, R. M. Vergara vs. State of California Tentative Decision (2014).

Appendix B. Tables and Figures

Table 1. Outcome Variable Summary Statistics

	All	21-Pt Pre-Screening Summ. Rating	60-Pt Screening Summ. Rating	Interview	Hired/ Offered	Hired Elsewhere
Total Obs (Teacher/Yr.)	4217	3944	1709	1238	538	498
Total Proportions	1.00	0.94	0.41	0.29	0.13	0.12
Applicant Information						
Certificated Employment Experience in Year Applied						
No Experience	0.83	0.84	0.68	0.63	0.49	0.53
	(0.38)	(0.36)	(0.47)	(0.48)	(0.50)	(0.50)
SPS District	0.11	0.09	0.22	0.28	0.43	0.03
	(0.31)	(0.28)	(0.42)	(0.45)	(0.49)	(0.17)
Other District	0.07	0.07	0.09	0.09	0.08	0.44
	(0.25)	(0.25)	(0.29)	(0.29)	(0.27)	(0.50)
Calculated Experience	3.18	3.23	3.87	3.73	3.24	4.43
	(4.66)	(4.64)	(5.02)	(4.74)	(4.23)	(5.30)
Student Teaching in SPS? (Y/N)	0.36	0.37	0.40	0.42	0.47	0.29
	(0.48)	(0.48)	(0.49)	(0.49)	(0.50)	(0.46)
21-Point Pre-Screening Rubric Summative Rating	NA	16.10	16.99	17.13	17.27	16.49
		(2.36)	(2.21)	(2.19)	(2.16)	(2.22)
60-Point Screening Rubric Summative Rating	NA	NA	41.34	43.62	45.66	40.19
			(7.32)	(6.19)	(5.74)	(7.06)
WESTB Average	-0.07	-0.07	-0.03	-0.02	0.02	-0.04
(Standardized statewide)	(0.75)	(0.75)	(0.75)	(0.75)	(0.70)	(0.75)
(N = 1364 Teachers)						
Outcomes*						
Value-Added						
Math	-0.05	-0.06	-0.03	-0.03	-0.01	-0.08
(N=348 Teacher/Yr)	(0.21)	(0.21)	(0.21)	(0.21)	(0.21)	(0.19)
Reading	-0.08	-0.09	-0.08	-0.07	-0.07	-0.09
(N=364 Teacher/Yr)	(0.17)	(0.17)	(0.17)	(0.18)	(0.18)	(0.17)
Absences (N=1057 Teacher/Yr)						
Total Annual Absences	6.92	6.62	7.38	7.51	7.27	5.28
	(5.35)	(5.09)	(5.24)	(5.32)	(5.33)	(5.10)
Total Monday/Friday Absences	3.12	2.98	3.33	3.37	3.29	2.44
	(2.50)	(2.42)	(2.51)	(2.50)	(2.48)	(2.47)
Attrit within 1 Year (N=1020 Teacher/Yr)						
School	0.46	0.46	0.46	0.44	0.40	0.47
	(0.50)	(.50)	(0.50)	(0.50)	(0.49)	(0.50)
District	0.30	0.31	0.29	0.27	0.20	0.39
	(0.46)	(0.46)	(0.45)	(0.44)	(0.40)	(0.49)
K-12 WA Public Schools	0.22	0.23	0.22	0.21	0.17	0.17
	(0.41)	(0.42)	(0.41)	(0.41)	(0.37)	(0.38)

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) on page 12. 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). *Observation numbers in the Outcomes panel represent the number of applications (at the teacher/year level) with associated outcome data. The numbers of observations of observed teacher/year outcome data are smaller, and are shown (conditional on having observed screening scores) as the number of clusters in each regression in **Tables 6-8**.

Table 2. Predictors of Student Achievement in Math and Reading

		Math		Reading	
(1) 21-Point Pre-Screening Rubric Summative Rating		0.044*	R^2= 0.638	0.016	R^2=0.595
		(0.022)	N = 4,797 (180)	(0.019)	N=4,695 (185)
(2) 21-Point Rubric Components	Experience	0.023	R^2= 0.650	0.042*	R^2=0.607
		(0.024)	N = 3,467 (124)	(0.022)	N = 2,774 (120)
	Skills	0.037	R^2=0.651	0.003	R^2=0.605
		(0.025)	N = 3,467 (124)	(0.019)	N = 2,774 (120)
	Recommendations	0.046*	R^2=0.652	0.054*	R^2=0.607
		(0.028)	N = 3,467 (124)	(0.028)	N = 2,774 (120)
(3) 60-Point Screening Rubric Summative Rating		0.081**	R^2=0.640	0.055*	R^2=0.596
		(0.029)	N = 3,407 (131)	(0.033)	N = 3,357 (135)
(4) 60-Point Rubric Components	Certificate & Education	-0.044	R^2=0.637	-0.039	R^2=0.595
		(0.040)	N = 3,407 (131)	(0.040)	N = 3,357 (135)
	Training	0.062*	R^2=0.638	0.056	R^2=0.596
		(0.034)	N = 3,407 (131)	(0.035)	N = 3,357 (135)
	Experience	0.097**	R^2=0.640	0.042	R^2=0.596
		(0.044)	N = 3,407 (131)	(0.033)	N = 3,357 (135)
	Classroom Management	0.084**	R^2=0.639	0.037	R^2=0.595
		(0.030)	N = 3,407 (131)	(0.029)	N = 3,357 (135)
	Flexibility	0.087**	R^2=0.639	0.072**	R^2=0.597
		(0.032)	N = 3,407 (131)	(0.030)	N = 3,357 (135)
	Instructional Skills	0.109**	R^2=0.641	0.061**	R^2=0.596
		(0.032)	N = 3,407 (131)	(0.027)	N = 3,357 (135)
	Interpersonal Skills	0.048	R^2=0.637	0.002	R^2=0.595
		(0.041)	N = 3,407 (131)	(0.028)	N = 3,357 (135)
	Cultural Competency	0.079**	R^2=0.639	0.016	R^2=0.595
		(0.034)	N = 3,407 (131)	(0.027)	N = 3,357 (135)
	Preferred Qualifications	0.051	R^2=0.637	0.075**	R^2=0.596
		(0.039)	N = 3,407 (131)	(0.035)	N = 3,357 (135)
	Letters of Recommendation	0.003	R^2=0.636	-0.087**	R^2=0.597
		(0.045)	N = 3,407 (131)	(0.038)	N = 3,357 (135)
(5) 21 and 60 Point Screening Rubric Summative Rating					
21-Point Rating		0.050		0.042	
		(0.033)	R^2=0.644	(0.029)	R^2=0.600
60-Point Rating		0.061*	N = 3,040 (112)	0.034	N = 2,892 (115)
		(0.036)		(0.042)	

Notes: Each of the specifications includes controls for prior student test scores in math and reading, a vector of student-level controls (gender, ethnicity, learning disability status, gifted program status, and free-or-reduced-lunch status), an indicator for currently working in Spokane, grade level effects, and year effects. Standard errors are clustered at the teacher/year level. The number of clusters in each analysis is presented in parentheses next to the total number of observations. *** p < .01, ** p < 0.05, * p < 0.10

Table 3: Predictors of Teacher Absences (Sick Days Taken)

		Average Yearly Absences		Monday and Friday Absences	
(1) 21-Point Pre-Screening Rubric Summative Rating		0.370	R^2= 0.083	0.235	R^2=0.069
		(0.310)	N = 453 (335)	(0.151)	N = 453 (335)
(2) 21-Point Rubric Components	Experience	0.242	R^2= 0.074	0.067	R^2=0.087
		(0.333)	N = 304 (231)	(0.164)	N = 304 (231)
	Skills	-0.333	R^2= 0.074	-0.114	R^2=0.088
		(0.454)	N = 304 (231)	(0.208)	N = 304 (231)
	Recommendations	-0.078	R^2= 0.072	-0.017	R^2=0.087
		(0.369)	N = 304 (231)	(0.169)	N = 304 (231)
(3) 60-Point Screening Rubric Summative Rating		-0.027	R^2=0.075	0.054	R^2=0.060
		(0.522)	N = 287 (213)	(0.284)	N = 287 (213)
(4) 60-Point Rubric Components	Certificate & Education	0.400	R^2=0.077	0.433*	R^2=0.071
		(0.542)	N = 287 (213)	(0.259)	N = 287 (213)
	Training	0.171	R^2=0.075	0.041	R^2=0.060
		(0.540)	N = 287 (213)	(0.315)	N = 287 (213)
	Experience	1.129**	R^2=0.093	0.355	R^2=0.067
		(0.454)	N = 287 (213)	(0.273)	N = 287 (213)
	Classroom Management	-0.203	R^2=0.076	-0.017	R^2=0.060
		(0.478)	N = 287 (213)	(0.225)	N = 287 (213)
	Flexibility	-0.079	R^2=0.075	-0.064	R^2=0.060
		(0.665)	N = 287 (213)	(0.338)	N = 287 (213)
	Instructional Skills	-0.371	R^2=0.077	-0.046	R^2=0.060
		(0.568)	N = 287 (213)	(0.254)	N = 287 (213)
	Interpersonal Skills	-0.515	R^2=0.079	-0.099	R^2=0.061
		(0.472)	N = 287 (213)	(0.258)	N = 287 (213)
	Cultural Competency	0.029	R^2=0.075	-0.244	R^2=0.064
		(0.482)	N = 287 (213)	(0.238)	N = 287 (213)
	Preferred Qualifications	0.339	R^2=0.077	0.301	R^2=0.065
		(0.635)	N = 287 (213)	(0.292)	N = 287 (213)
	Letters of Recommendation	-0.186	R^2=0.075	-0.070	R^2=0.061
		(0.422)	N = 287 (213)	(0.263)	N = 287 (213)
(5) 21 and 60 Point Screening Rubric Summative Rating					
21-Point Rating		0.270		0.132	
		(0.784)	R^2=0.074	(0.345)	R^2=0.059
60-Point Rating		-0.074	N = 272 (205)	-0.004	N = 272 (205)
		(0.543)		(0.298)	

Notes: Each specification controls for gender, ethnicity, school size, school percentages for students eligible for free/reduced lunch and for under-represented minorities, and indicators for currently working in Spokane, school level and Title I status. Standard errors are clustered at the teacher level. Clusters are presented in parentheses. *** $p < .01$, ** $p < 0.05$, * $p < 0.10$

Table 4. Predictors of Teacher Attrition from School, District, and State

		School		District		WA K-12 Public Schools	
(1) 21-Point Pre-Screening Rubric Summative Rating		-0.219***	R^2= 0.0636	-0.158*	R^2=0.0638	-0.166	R^2=0.0572
		(0.078)	N = 1,148 (739)	(0.093)	N=1,229 (741)	(0.113)	N=1,291 (749)
(2) 21-Point Rubric Components	Experience	-0.114	R^2= 0.0773	-0.067	R^2=0.0802	-0.015	R^2=0.0890
		(0.087)	N = 814 (552)	(0.107)	N = 865 (548)	(0.131)	N=896 (551)
	Skills	-0.227**	R^2=0.0814	-0.192*	R^2=0.0840	-0.208	R^2=0.0939
		(0.095)	N = 814 (552)	(0.109)	N = 865 (548)	(0.138)	N=896 (551)
	Recommendations	-0.192**	R^2=0.0799	-0.156	R^2= 0.0822	-0.159	R^2=0.0916
		(0.091)	N = 814 (552)	(0.115)	N = 865 (548)	(0.137)	N=896 (551)
(3) 60-Point Screening Rubric Summative Rating		-0.271***	R^2=0.0670	-0.259**	R^2=.0703	-0.307***	R^2=0.0608
		(0.090)	N = 1,194 (744)	(0.104)	N=1,278 (747)	(0.118)	N=1,339 (761)
(4) 60-Point Rubric Components	Certificate & Education	0.028	R^2=0.0598	0.024	R^2=0.0653	0.077	R^2=0.0533
		(0.095)	N = 1,194 (744)	(0.111)	N=1,278 (747)	(0.127)	N=1,339 (761)
	Training	-0.163*	R^2=0.0624	-0.153	R^2=0.0674	-0.186	R^2=0.0558
		(0.091)	N = 1,194 (744)	(0.106)	N=1,278 (747)	(0.126)	N=1,339 (761)
	Experience	-0.251***	R^2=0.0663	-0.247**	R^2=0.0710	-0.239**	R^2=0.0579
		(0.088)	N = 1,194 (744)	(0.102)	N=1,278 (747)	(0.118)	N=1,339 (761)
	Classroom Management	-0.241***	R^2=0.0665	-0.221**	R^2=0.0704	-0.295***	R^2=0.0617
		(0.086)	N = 1,194 (744)	(0.097)	N=1,278 (747)	(0.102)	N=1,339 (761)
	Flexibility	-0.221**	R^2=0.0647	-0.224**	R^2=0.0700	-0.260**	R^2=0.0589
		(0.090)	N = 1,194 (744)	(0.103)	N=1,278 (747)	(0.115)	N=1,339 (761)
	Instructional Skills	-0.228**	R^2=0.0650	-0.272***	R^2=0.0721	-0.366***	R^2=0.0648
		(0.089)	N = 1,194 (744)	(0.105)	N=1,278 (747)	(0.116)	N=1,339 (761)
	Interpersonal Skills	-0.275***	R^2=0.0670	-0.339***	R^2=0.0760	-0.348***	R^2=0.0630
		(0.091)	N = 1,194 (744)	(0.102)	N=1,278 (747)	(0.120)	N=1,339 (761)
	Cultural Competency	-0.150*	R^2=0.0618	-0.104	R^2=0.0661	-0.225**	R^2=0.0570
		(0.090)	N = 1,194 (744)	(0.103)	N=1,278 (747)	(0.114)	N=1,339 (761)
	Preferred Qualifications	-0.168*	R^2=0.0624	-0.218**	R^2=0.0691	-0.255**	R^2=0.0579
		(0.090)	N = 1,194 (744)	(0.106)	N=1,278 (747)	(0.118)	N=1,339 (761)
	Letters of Recommendation	-0.203*	R^2=0.0629	-0.078	R^2=0.0656	0.067	R^2=0.0532
		(0.104)	N = 1,194 (744)	(0.116)	N=1,278 (747)	(0.125)	N=1,339 (761)
(5) 21 and 60 Point Screening Rubric Summative Rating							
21-Point Rating		-0.257***		-0.215**		-0.211*	
		(0.088)	R^2=0.0802	(0.103)	R^2=0.0716	(0.126)	R^2=0.0700
60-Point Rating		-0.251**	N = 1,032 (665)	-0.224**	N=1,102 (669)	-0.287**	N=1,160 (679)
		(0.101)		(0.113)		(0.135)	

Notes: Each of the specifications includes controls for gender, ethnicity, school size, school percentages for students eligible for free or reduced lunch, school percentages for under-represented minorities, and indicators for currently working in Spokane, school level and Title I status. Standard errors are clustered at the teacher level. The number of clusters in each analysis is presented in parentheses next to the total number of observations. *** $p < .01$, ** $p < 0.05$, * $p < 0.10$

Table 5. The Effect of Screening Scores on Outcomes, With and Without Selection Correction

	Math		Reading	
VARIABLES	(1)	(2)	(3)	(4)
21-Pt Screen	-0.033	-0.029	0.011	0.029
	(0.056)	(0.054)	(0.038)	(0.041)
60-Pt Screen	0.078	0.079	0.014	0.004
	(0.074)	(0.076)	(0.063)	(0.067)
Mills Ratio (λ)		-0.075		-0.212
		(0.122)		(0.146)
Observations	1,519 (64)		1,425 (62)	
R-Squared	0.610		0.592	
Overidentification p-value		0.934		0.323
	Absences		Monday/Friday Absences	
21-Pt Screen	1.300*	1.276*	0.157	0.115
	(0.736)	(0.759)	(0.326)	(0.337)
60-Pt Screen	-0.242	-0.311	0.133	0.009
	(0.684)	(0.699)	(0.431)	(0.433)
Mills Ratio (λ)		-0.591		-1.052
		(1.685)		(0.815)
Observations	140 (106)		140 (106)	
R-Squared	0.228		0.199	
Overidentification p-value		0.887		0.901
	1-Year District Attrition		3-Year District Attrition	
21-Pt Screen	0.041	0.041	0.042	0.050
	(0.035)	(0.038)	(0.042)	(0.044)
60-Pt Screen	-0.089*	-0.089*	-0.120*	-0.112*
	(0.052)	(0.053)	(0.064)	(0.065)
Mills Ratio (λ)		-0.001		0.087
		(0.108)		(0.145)
Observations	197		141	
R-Squared	0.046		0.093	
Overidentification p-value		0.545		0.614

Estimates are produced using standard models as presented in **Tables 2-4**, except that the sample is limited to those hired into Spokane in the sampling window, and the selection correction is included in models (2) and (4). The number of clusters in each analysis is presented in parentheses next to the total number of observations. *** $p < .01$, ** $p < 0.05$, * $p < 0.10$