Can a Commercial Screening Tool Help Select Better Teachers?

Olivia L. Chi
Boston University

Matthew A. Lenard iD
Harvard University

Improving teacher selection is an important strategy for strengthening the quality of the teacher workforce. As districts adopt commercial teacher screening tools, evidence is needed to understand these tools' predictive validity. We examine the relationship between Frontline Education's TeacherFit instrument and newly hired teachers' outcomes. We find that a 1 SD increase on an index of TeacherFit scores is associated with a 0.06 SD increase in evaluation scores. However, we also find evidence that teachers with higher TeacherFit scores are more likely to leave their hiring schools the following year. Our results suggest that TeacherFit is not necessarily a substitute for more rigorous screening processes that are conducted by human resources officials, such as those documented in recent studies.

Keywords: *school/teacher effectiveness, teacher characteristics, educational policy, econometric analysis*

Introduction

An extensive literature documents substantial variation in teacher quality, and evidence shows that teacher quality impacts students' long-run outcomes (see Jackson et al., 2014 for a review). An important strategy for strengthening the quality of the teacher workforce is to identify and hire the applicants who will be most effective in the classroom. Recent studies suggest that information gathered during a district's application-screening process could be compiled to make predictions about teacher effectiveness. Goldhaber et al. (2017) find that scores from teacher selection rubrics used to rate applicants in Spokane Public Schools predict teacher value-added to test scores and teacher retention. Using data from Washington, D.C., Jacob et al. (2018) find that applicants' background measures (e.g.,

undergraduate GPA) and scores on screening measures are predictive of teachers' evaluation scores. Bruno and Strunk (2019) use data from the Los Angeles Unified School District to show that scores from the district office's standardized screening system are predictive of teacher impacts on test scores, evaluation scores, and attendance. Sajjadiani et al. (2019), applying machine learning techniques to data from the Minneapolis Public School District, find that the relevance of applicants' work experience and their attributions for leaving past jobs predict teacher performance (student evaluations, observation scores, and value-added to test scores) and turnover.

These recent studies demonstrate promising ways in which information from the time of application can be harnessed to select better teachers. However, in these studies, districts'

central office human resources officials play a large role in prescreening and scoring applications. This systematic collection and scoring of applicant information for use in hiring decisions may be a barrier in some school districts. Prior research suggests that district central offices vary in how they support schools in recruiting and screening teacher applicants, and principals vary in the degree to which they use data in their hiring decisions (Cannata et al., 2017). The teacher hiring process may even be rushed and information-poor (Liu & Johnson, 2006).

To overcome this limitation, school districts across the United States are turning to the private sector for new teacher screening tools that claim to use "big data" to help identify better teachers from the pool of teacher applicants (Flanigan, 2016; Simon, 2014). Commercial screening instruments have existed for decades, such as the Haberman Star Teacher PreScreener and Gallup's Teacher Perceiver Interview, with limited evidence suggesting that scores from these two tools are modestly related to teacher performance (Metzger & Wu, 2008; Rockoff et al., 2011; see Supplementary Appendix A in the online version of the journal for additional background). However, additional commercial screening instruments have arrived on the market, boasting data-driven screening scores (Simon, 2014). Many districts currently pay firms to use these screening instruments, which typically include assessments that applicants take while completing online teacher job applications (Simon, 2014). Despite the growing popularity of such commercial screening tools, there exists limited evidence on whether these tools are effective in predicting teacher performance.

To extend this literature, we study the extent to which applicants' scores from a "big data" commercial screening tool, Frontline Education's TeacherFit instrument, predict teacher outcomes in a large U.S. school district. Specifically, we ask: Are the results from the TeacherFit screening tool predictive of teachers' evaluation scores, absences, retention in the same school/district, and impacts on student test scores?

We use data from the Wake County Public School System (WCPSS) in North Carolina, the 14th largest school district in the nation (de Brey et al., 2021). Beginning in January 2016, WCPSS required teacher applicants to take the TeacherFit

assessment to submit their applications. We study the predictive validity of applicants' TeacherFit scores among the new teacher applicant pool for WCPSS in school years 2016–2017 and 2017–2018. We find that a one standard deviation increase on an index of TeacherFit scores is associated with a 0.06 standard deviation increase in the evaluation scores that teachers receive from principals. In addition, we find evidence that teachers with higher TeacherFit scores are more likely to leave their hiring schools after the first year. We do not find a significant relationship between TeacherFit scores and either value-added to math or English Language Arts (ELA) test scores. To alleviate concerns of bias from sample selection, we estimate a Heckman selection model and find that sample selection–corrected estimates are similar to our estimates without corrections.

Setting, Data, and Measures

Frontline Education is a school administration software provider whose broad portfolio of products reaches over 10,000 clients (Frontline Education, 2021). WCPSS began using Frontline's Web-based recruiting platform, Frontline Recruiting and Hiring, to manage job postings and applications in the summer of 2015. Beginning in January 2016, WCPSS required teacher applicants to take Frontline's screening assessment, TeacherFit, to submit their applications via the district's site on Frontline's online platform. Frontline Education's website states that their assessments are "driven by university backed research" and "leverage 'machine learning'" (Grunwell, 2016). According to a sales webinar, the assessment creators developed the items based on interviews with subject matter experts, analysis of job descriptions, and reviews of research, followed by testing of the items on teacher and school employees (Reese, 2018). "Hundreds of thousands" of applicants have taken the assessment since it became available in 2008 (Frontline Education, 2022). However, to our knowledge, the extent to which scores from the TeacherFit instrument can identify effective teachers has not yet been documented in peer-reviewed literature.

The TeacherFit assessment, which claims to help "identify outstanding teachers" (Grunwell, 2016), takes approximately 20 to 30 minutes,

does not allow for blank responses, and does not have obvious "correct" or "incorrect" answers. Rather, the items attempt to assess applicants' attitudes, beliefs, habits, and personality traits by requiring applicants to address situational prompts and attitudinal statements by selecting Likert-type scale responses. Following the assessment, Frontline constructs the applicants' scores and makes them available to WCPSS administrators and school principals on the online hiring dashboard. Each candidate receives an overall score and separate scores on each of the six dimensions: Fairness & Respect, Concern for Student Learning, Adaptability, Communication & Persuasion, Planning & Organizing, and Cultural Competence. Each score falls on a 1-to-9 scale. (See Supplementary Appendix A in the online version of the journal for additional background on TeacherFit.)

While applicants' scores were made available to administrators on the hiring platform, school principals did not receive strict or explicit guidance from the WCPSS Human Resources office on how to use or interpret the scores, keeping with a tradition of a decentralized hiring process in WCPSS. While principals received communication that TeacherFit scores could help identify strong candidates, they also received messaging that the TeacherFit assessment is only one part of the hiring process, and they were free to pursue candidates that do not score well. (See Supplementary Appendix B in the online version of the journal for additional background on teacher hiring.)

We use application and administrative data to study the new teacher applicant pool for school years 2016–2017 and 2017–2018. We link teacher applicants' scores and application information to WCPSS administrative data, which includes teacher characteristics, assignments, evaluation scores, absences, and links to students. Individuals who are not previously observed as WCPSS teachers are included in the new teacher applicant pool if they (a) submitted teacher applications in the calendar years 2016 and 2017 and/or (b) are newly hired teachers in 2016–2017 and 2017–2018. The applicant pool includes 12,548 individuals, of whom 2,367 are observed as newly hired teachers in either 2016–2017 or 2017–2018. However, TeacherFit scores are missing for 8%

of the applicant pool. Therefore, 11,491 individuals, of whom 2,104 are observed as new hires across 184 schools, are eligible for inclusion in the analyses below.

Table 1 provides summary statistics of the TeacherFit scores for the individuals in the teacher applicant pool with non-missing scores ($N = 11,491$). The TeacherFit overall score (1–9 scale) has a mean of 6.04 ($SD = 1.80$), while the means of the scores on the six dimensions range from 5.23 ($SD = 2.08$) in Cultural Competence to 6.12 ($SD = 1.82$) in Adaptability. In the analyses below, we use a TeacherFit index score, which we construct by summing the scores for the six dimensions and then standardizing to have a mean of 0 and unit standard deviation.

Teachers' Outcome Measures

Our primary outcome of interest is teachers' evaluation scores. In the North Carolina Teacher Evaluation Process, teachers must be reviewed annually by their principals or a similar designated evaluator. To construct teachers' annual evaluation scores, we fit a Graded Response Model (GRM) on the ratings that teachers receive on each element of their Summary Rating Forms from the North Carolina evaluation regime (Kraft et al., 2020). GRM models are in the family of Item Response Theory (IRT) models that are commonly used in educational and psychological assessment. GRMs are developed for ordered categorical items, such as the five-category scale on the Summary Rating Forms in the NC Teacher Evaluation Process (Samejima, 1968). We then standardize the GRM scores within-year to have a mean of 0 and a standard deviation of 1.

We also examine additional teacher outcome measures, including the number of days a teacher is absent, retention in the same school and same district in the following year, and impacts on math and ELA test scores, when available. Retention in the same school (district) is a binary indicator of whether a teacher returns to teach in the same school (in WCPSS) in the following school year. By definition, those who remain teaching in the same school in the following year also remain teaching in the same district, WCPSS, in the following year. However, those who remain teaching in WCPSS in the following

TABLE 1

Summary Statistics of Scores for the New Teacher Applicant Pool

	New teacher applicant pool				New hires				Non-hires				Difference
	Mean	SD	Min.	Max.	Mean	SD	Min.	Max.	Mean	SD	Min.	Max.	New hires − non-hires
TeacherFit Overall Score	6.04	1.80	1.00	9.00	6.21	1.69	1.00	9.00	6.01	1.82	1.00	9.00	0.21***
Dimensions													
Fairness and respect	5.56	1.84	1.00	9.00	5.79	1.75	1.00	9.00	5.51	1.86	1.00	9.00	0.28***
Concern for student learning	5.95	1.86	1.00	9.00	6.17	1.74	1.00	9.00	5.91	1.88	1.00	9.00	0.26***
Adaptability	6.12	1.82	1.00	9.00	6.13	1.73	1.00	9.00	6.11	1.84	1.00	9.00	0.01
Communication and persuasion	6.05	1.85	1.00	9.00	6.06	1.80	1.00	9.00	6.05	1.86	1.00	9.00	0.01
Planning and organizing	6.11	1.83	1.00	9.00	6.23	1.75	1.00	9.00	6.09	1.84	1.00	9.00	0.15**
Cultural competence	5.23	2.08	1.00	9.00	5.48	2.01	1.00	9.00	5.17	2.09	1.00	9.00	0.30***
Composite measures													
TeacherFit Sum Score (raw)	35.03	8.48	6.00	54.00	35.85	8.02	7.00	54.00	34.84	8.57	6.00	54.00	1.01***
Std. TeacherFit Sum Score	0.07	0.98	−3.27	2.26	0.17	0.92	−3.16	2.26	0.05	0.99	−3.27	2.26	0.12***
Missing TeacherFit Score	0.08				0.11				0.08				0.03***
N (applicants with non-missing scores)	11,491				2,104				9,387				11,491
N (teachers)	12,548				2,367				10,181				12,548

Notes. The new teacher applicant pool includes 12,548 individuals who submitted teacher applications in the calendar years 2016 and 2017 and/or are newly hired teachers in 2016–2017 and 2017–2018.

**$p < .01$. ***$p < .001$.

533

year are not necessarily teaching in the same school in the following year. To calculate teachers' impacts on test scores, we estimate value-added models for math and ELA teachers of fourth- through eighth-grade students (see Supplementary Appendix C in the online version of the journal).

Empirical Strategy

To estimate whether TeacherFit scores are predictive of teacher-level outcomes, we use ordinary least squares to estimate

$$Y_{jkt} = \delta_0 + \delta_1 \text{Score}_j + \\ T_j\delta_2 + S_{jkt}\delta_3 + \delta_t + \delta_h + v_{jkt}, \quad (1)$$

where Y_{jkt} indicates the outcome of teacher j in school k at time t. Score_j refers to teacher j's standardized TeacherFit index scores. The coefficient of interest δ_1 is the expected change in the outcome Y associated with a one standard deviation increase in an applicant's TeacherFit index score. T_j represents a vector of indicators for the teacher characteristics of race, gender, and experience. In theory, including controls for teacher experience may account for variation in the outcome that would instead be attributable to differences in TeacherFit scores in the absence of experience controls. However, we include these to address whether and to what extent TeacherFit scores can provide additional predictive information, above and beyond what is already known from resumes at the time of hiring, such as teacher experience. S_{jkt} represents a vector of annual school characteristics of teacher j's school k, including student gender, race, Limited English Proficient (LEP) status, and special education status, aggregated to school level, along with mean school-level prior test scores and school size. δ_t represents year indicators, and δ_h are indicators for the number of years since being newly hired. Standard errors are clustered at the teacher level.

To alleviate concerns of bias stemming from the possibility that teachers with higher (or lower) TeacherFit scores are systematically sorting into schools or job assignments (e.g., third grade, middle/secondary math, middle/secondary ELA) that enable teachers to have

better performance measures, we also estimate these models with (a) school fixed effects in place of school-level characteristics, and (b) job assignment fixed effects. In these models, the identifying variation comes from applicants who are hired into the same school and applicants who are hired into the same job assignment.

A limitation worth noting is the potential for bias stemming from the possibility that administrators may—subconsciously or otherwise—reward higher TeacherFit scores, which they observed during the hiring process, with higher evaluation scores. This could bias our estimate of the relationship between TeacherFit and evaluation scores upward. However, we suspect that this would not be a large source of bias in this specific context as administrators were not advised or guided to put much stock in TeacherFit scores.

Sample Selection Correction

Although we can observe the outcomes of interest for hired applicants, we lack information on how non-hired applicants would have performed had they been hired. In other words, we cannot examine the relationship between TeacherFit scores and teacher performance for the full range of applicants. Given that the individuals making hiring decisions in WCPSS are likely trying to select applicants whom they perceive to be of higher quality, the hiring process may introduce selection bias into our estimates. Specifically, we are concerned that low-scoring individuals who end up hired as new teachers in WCPSS, in spite of their low scores, are particularly impressive in ways that are (a) unobservable and (b) correlated with their performance as teachers. To alleviate concerns of bias from sample selection, similar to Goldhaber et al. (2017), we estimate sample selection–corrected models using a Heckman selection model (Heckman, 1979). We identify the model using a function of the school size growth among the set of schools to which applicants submit applications (see Supplementary Appendix D in the online version of the journal). The extent of school size growth among an applicant's set of schools is predictive of the likelihood they become a newly hired teacher in WCPSS but is unrelated to

on-the-job performance as a teacher. To preview, we find that sample selection–corrected estimates are similar to our estimates without corrections, alleviating concerns of large bias from sample selection.

Results

We present estimates from Equation 1 in Table 2. The odd-numbered columns include controls for school characteristics, while the even-numbered columns replace the school characteristics controls with school fixed effects and add job assignment fixed effects. As shown in Columns (1) and (2), we find that scores from the TeacherFit screening tool significantly predict teachers' standardized evaluation scores. A one standard deviation increase in the TeacherFit index is associated with a 0.08 *SD* increase in evaluation scores in our baseline model. After including school and job assignment fixed effects, this estimate attenuates slightly but remains statistically significant at 0.06 *SD*. This magnitude is about 16% of the estimated within-teacher returns to experience after 1 year of teaching (0.38 *SD*, estimated with WCPSS data from 2015–2016 through 2017–2018). Columns (3) and (4) examine the relationship between the TeacherFit scores and teacher absences. The coefficients are positive and small, but the results are not statistically significant—null results that are consistent with those reported by both Goldhaber et al. (2017) and Rockoff et al. (2011).

In columns (5) through (8), we examine the relationship between TeacherFit scores and teacher retention. Surprisingly, in our baseline models, we find that a one standard deviation increase in the TeacherFit index is associated with a 3.4 percentage point *decrease* in the likelihood of remaining as a teacher in the same school in the following year (column [5]) and a 2.4 percentage point *decrease* in the likelihood of remaining in WCPSS in the following year (column [7]). After including school and job assignment fixed effects, the results attenuate slightly. These retention results provide some evidence that TeacherFit scores are negatively associated with within-school and within-district teacher retention.

These findings are contrary to the retention results from Goldhaber et al. (2017), who find that higher scores on a screening rubric, which is completed by human resources hiring officials, predict an increase in district retention, as well as results from Jacob et al. (2018), who find that higher screening scores predict a higher likelihood of remaining in the hiring school. Jacob et al. (2018), however, do find that teachers with better academic background scores are more likely to leave their hiring school and more likely to leave District of Columbia Public Schools (DCPS) after their first year.

Columns (9) through (12) present results of the relationship between TeacherFit scores and value-added to math and ELA test scores of students in Grades 4 through 8, measured in student-level test score SDs. All our estimates are relatively close to 0 and are statistically insignificant. The point estimates for math value-added scores, 0.005 (column 9) and 0.004 (column 10) are equivalent to 0.025 *SD* and 0.020 *SD*, respectively, in teacher-level SDs. The point estimates for ELA value-added scores, 0.002 (column 11) and −0.002 (column 12) are equivalent to 0.014 *SD* and −0.014 *SD*, respectively, in teacher-level SDs.

Selection-Corrected Estimates

Table 3 presents our selection-corrected estimates of the relationship between TeacherFit scores and our outcomes of interest. Here, the sample of hired teachers is smaller than that included in Table 2, as it is limited to applicants who (a) submitted teacher applications in the same calendar year in which they are hired, and (b) apply to schools where we can measure the change in school size between the prior year and the time of application. We also only include newly hired teachers' first-year in the data, and the model is fit at the person-level. In this table, we present estimates from our baseline model (i.e., including school characteristics, and absent school and job assignment fixed effects), displaying results without and with the sample correction in the odd and even columns, respectively. The magnitude of the estimates appear substantively similar without and with the sample selection correction, alleviating concerns of large bias introduced by sample selection. Given these results,

TABLE 2

Relationship between TeacherFit Scores and Teacher Outcomes

	Evaluation scores		Absences		Stay in school		Stay in district		Math VA		ELA VA	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Std. TeacherFit Sum Score	0.077***	0.060***	0.059	0.226	−0.034***	−0.029**	−0.024**	−0.021*	0.005	0.004	0.002	−0.002
	(0.019)	(0.017)	(0.166)	(0.171)	(0.009)	(0.010)	(0.008)	(0.009)	(0.016)	(0.019)	(0.011)	(0.013)
School Characteristics	Y		Y		Y		Y		Y		Y	
School FE		Y		Y		Y		Y		Y		Y
Job Assignment FE		Y		Y		Y		Y		Y		Y
Unique Teachers	1,884	1,884	2,063	2,063	2,088	2,088	2,088	2,088	236	236	246	246
N (teacher-years)	2,580	2,580	2,830	2,830	2,873	2,873	2,873	2,873	302	302	316	316
N (teacher-years) w/ 2+ teachers												
in same school		2,578		2,828		2,871		2,871		282		292
in same job assignment		2,580		2,828		2,871		2,871		302		316

Notes. Clustered standard errors at the teacher-level are in parentheses. All regressions include a vector of indicators for the teacher characteristics of race, gender, and experience. School characteristics include student gender, race, Limited English Proficient (LEP) status, and special education status, aggregated to school-level, along with mean school-level prior test scores and school size. VA = value-added; FE = fixed effects.

*$p < .05$. **$p < .01$. ***$p < .001$.

TABLE 3

Sample Selection-Corrected Estimates

	Evaluation score		Absences		Stay in school		Stay in district		Math VA		ELA VA	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Std. TeacherFit Sum Score	0.079***	0.091**	0.204	0.209	−0.028*	−0.020	−0.027**	−0.042**	0.028†	0.018	0.015	0.027†
	(0.020)	(0.028)	(0.185)	(0.262)	(0.012)	(0.017)	(0.010)	(0.015)	(0.016)	(0.018)	(0.014)	(0.015)
Mills ratio		0.231		0.084		0.137		−0.246		−0.214		0.250
		(0.382)		(3.382)		(0.211)		(0.185)		(0.198)		(0.154)
Sample selection correction		Y		Y		Y		Y		Y		Y
N (teachers)	1,662	1,662	1,844	1,844	1,874	1,874	1,874	1,874	232	232	231	231
N (applicants first-stage)		10,706		10,701		10,701		10,701		10,713		10,714
F (excluded variable)		23.845		22.442		25.840		25.840		16.891		13.645

Notes. Clustered standard errors at the teacher-level are in parentheses. All regressions include a vector of indicators for the teacher characteristics of race, gender, and experience, as well as student gender, race, Limited English Proficient (LEP) status, and special education status, aggregated to school-level, along with mean school-level prior test scores and school size. VA = value-added.

†$p < .10.$ *$p < .05.$ **$p < .01.$ ***$p < .001.$

we prefer the estimates presented in Table 2 from our larger and less restrictive sample.

Conclusion

We find that scores from the TeacherFit instrument have some capacity to predict teacher performance as measured by evaluation scores from principals. We find that a one standard deviation increase on an index of TeacherFit scores is associated with a 0.06 standard deviation increase in evaluation scores. However, we do not find a significant relationship between TeacherFit scores and teacher impacts on test scores. Furthermore, we find some evidence that teachers with higher TeacherFit scores are more likely to leave their hiring schools after the first year.

These results suggest that the TeacherFit commercial screening tool is not necessarily a substitute for the promising screening processes that are conducted by human resources officials as described in the studies by Bruno and Strunk (2019), Goldhaber et al. (2017), and Jacob et al., (2018). The screening scores from these more elaborate screening processes appear to be stronger predictors of desirable teacher outcomes, and investing in these screening systems may have higher payoffs than investing in commercial screening tools that may be cheaper and easier to implement. These TeacherFit results are also more modest than those documented in Rockoff et al.'s (2011) examination of Haberman PreScreener scores, though the differences in teacher characteristics between studies (WCPSS new hires vs. New York City elementary/middle math teachers) make comparison difficult. Nevertheless, our results focus on just one example of a "big data" commercial screening instrument. As districts adopt and/or continue using commercial screening tools, researchers and practitioners should monitor the predictive validity of these tools to ensure that scores from these tools contain information that can be used to improve teacher selection and retention.

ORCID iD

Matthew A. Lenard https://orcid.org/0000-0003-2234-0666

References

Bruno, P., & Strunk, K. O. (2019). Making the cut: The effectiveness of teacher screening and hiring in the Los Angeles Unified School District. *Educational Evaluation and Policy Analysis*, *41*(4), 426–460. https://doi.org/10.3102/0162373719865561

Cannata, M., Rubin, M., Goldring, E., Grissom, J. A., Neumerski, C. M., Drake, T. A., & Schuermann, P. (2017). Using teacher effectiveness data for information-rich hiring. *Educational Administration Quarterly*, *53*(2), 180–222.

de Brey, C., Synder, T. D., Zhang, A., & Dillow, S. A. (2021). *Digest of education statistics 2019 (NCES 2021-009)*. National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Flanigan, R. L. (2016, January 25). More districts mine data to refine teacher hiring. *Education Week*. https://www.edweek.org/

Frontline Education. (2021, December 27). *Frontline education recognized for adding value to K-12 administrators and the EdTech industry by a variety of organizations in 2021* [Press release]. https://www.frontlineeducation.com/news/frontline-education%E2%80%AFrecognized-for-adding%E2%80%AFvalue%E2%80%AFto-k-12-administrators-and-the-edtech-industry-by-a-variety-of-organizations-in-2021/

Frontline Education. (2022). *Application screening assessments*. https://www.frontlineeducation.com/solutions/recruiting-hiring/screening-assessments/

Goldhaber, D., Grout, C., & Huntington-Klein, N. (2017). Screen twice, cut once: Assessing the predictive validity of applicant selection tools. *Education Finance and Policy*, *12*(2), 197–223. https://doi.org/10.1162/EDFP_a_00200

Grunwell, A. (2016, August 1). *Applicant screening assessments: Frequently asked questions*. https://www.frontlineeducation.com/blog/applicant-screening-assessments-faqs/

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, *47*(1), 153. https://doi.org/10.2307/1912352

Jackson, C. K., Rockoff, J. E., & Staiger, D. O. (2014). Teacher effects and teacher-related policies. *Annual Review of Economics*, *6*(1), 801–825. https://doi.org/10.1146/annurev-economics-080213-040845

Jacob, B. A., Rockoff, J. E., Taylor, E. S., Lindy, B., & Rosen, R. (2018). Teacher applicant hiring and teacher performance: Evidence from DC public schools. *Journal of Public Economics*, *166*, 81–97. https://doi.org/10.1016/j.jpubeco.2018.08.011

Kraft, M. A., Papay, J. P., & Chi, O. L. (2020). Teacher skill development: Evidence from performance ratings by principals. *Journal of Policy Analysis and Management*, *39*(2), 315–347. https://doi.org/10.1002/pam.22193

Liu, E., & Johnson, S. M. (2006). New teachers' experiences of hiring: Late, rushed, and information-poor. *Educational Administration Quarterly*, *42*(3), 324–360. https://doi.org/10.1177/0013161X05282610

Metzger, S. A., & Wu, M. J. (2008). Commercial teacher selection instruments: The validity of selecting teachers through beliefs, attitudes, and values. *Review of Educational Research*, *78*(4), 921–940.

Reese, S. (2018). *Applitrack selection assessments* [Webinar]. Frontline Technologies. https://www.youtube.com/watch?v=Tl7ngKG1xoI

Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2011). Can you recognize an effective teacher when you recruit one? *Education Finance and Policy*, *6*(1), 43–74.

Sajjadiani, S., Sojourner, A. J., Kammeyer-Mueller, J. D., & Mykerezi, E. (2019). Using machine learning to translate applicant work history into predictors of performance and turnover. *Journal of Applied Psychology*, *104*(10), 1207–1225. https://doi.org/10.1037/apl0000405

Samejima, F. (1968). Estimation of latent ability using a response pattern of graded scores. *ETS Research Bulletin Series*, *1968*(1), i–169. https://doi.org/10.1002/j.2333-8504.1968.tb00153.x

Simon, S. (2014, December 29). Teacher hopefuls go through big data wringer. *Politico*. https://www.politico.com

Authors

OLIVIA L. CHI, PhD, is an assistant professor at Boston University's Wheelock College of Education and Human Development. Her research focuses on the economics of education, teacher labor markets, and measures of teacher quality.

MATTHEW A. LENARD, MA is a PhD candidate at the Harvard Graduate School of Education. His research focuses on the causal impacts of K–12 education programs and policies.