

The power of teacher selection to improve education

Brian A. Jacob

Executive Summary

This report describes the findings from a new study of the teacher selection process in Washington, DC public schools. In 2009, the district created a centralized application process to streamline hiring by screening out less desirable candidates. Following collection of standard information, applicants are asked to complete up to three additional in-depth assessments: (1) a subject-specific written assessment to assess their understanding of content and instructional practices; (2) a 30-minute structured interview with district personnel; and (3) a teaching “audition” that district personnel observed and evaluated.

We find that each of these measures—along with several more traditional indicators of academic achievement (e.g., undergraduate GPA)—strongly predict an individual’s performance on the district’s teacher evaluation system. Indeed, candidates in the top quartile of applicant quality score 0.6 standard deviations above applicants in the bottom quartile, a difference that is equivalent to the improvement that an average teacher makes between her first and third year on the job. These results suggest that improving teacher selection can be a relatively low cost way to improve the teacher workforce.

At the same time, we find that there still exists considerable scope for improving teacher selection in the district. Many candidates who were not hired had application scores exceeding the average of those who were hired. And only 30 percent of the very best applicants ended up working in the district. We conclude that having a sophisticated application system is not sufficient. Districts also must pay careful attention to how principals access and utilize the information generated by the system.

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

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

Over the past two decades, researchers have confirmed what many children and parents always believed—the individual classroom teacher is the key to a successful education. Research has documented the vast difference in student learning between classrooms with an effective teacher and those with an ineffective one. More interestingly, studies indicate that the most common measures of quality—e.g., certification, advanced degrees, and long tenure—are weak predictors of which teachers are most effective.

Policymakers have responded to these findings by focusing education reform on teacher effectiveness. Most states have mandated comprehensive new teacher evaluation systems that attempt to assess a teacher’s contribution to student learning. And many districts now use the results of such evaluations to dismiss poorly performing teachers and recognize exemplary teachers.

While there is some evidence that new evaluation systems have improved the quality of teaching in some districts,ⁱ they have also faced considerable hurdles. Critics argue that many of the evaluations are not particularly reliable, and might even systematically overstate or understate a teacher’s effectiveness depending on the caliber of students in his or her class. Finally, even when implemented in a thoughtful and responsible way, teacher evaluations can be quite costly.

Another popular way to improve instruction is through professional development. Unfortunately, it has proven very difficult to implement effective professional development programs at scale.ⁱⁱ The same can be said about preservice teacher education.ⁱⁱⁱ

But there is one approach that has not received much attention from the research or policy community—namely, improving the initial hiring of teachers. This idea is not new, as evidenced by the epigraph above. The pervasive belief that there is a perennial shortage of teachers may explain why this approach is so often overlooked. If there are few applicants to choose from, there is little need to worry about which candidate is chosen. But, contrary to this common view, many districts have an abundance of certified applicants in

many fields.^{iv}

Another potential reason that this approach has not received much attention is that teacher hiring is highly decentralized, with schools and principals creating their own idiosyncratic policies. The lack of a single centralized system makes it difficult not only to collect data on applicant characteristics, but also to determine which characteristics should be studied.

Improving selection has the potential to improve the teacher workforce at relatively low cost. It not only circumvents the political and financial costs of dismissing poorly performing teachers, but also decreases students’ exposure to ineffective instruction. And while the collection of performance data on teachers can be quite costly and often entails difficult labor negotiations, schools and school districts have considerable latitude to require applicants to submit information as part of the hiring process.

In a recently released study, my colleagues and I examined the teacher selection process in the District of Columbia Public Schools (hereafter DCPS) to learn how various measures might be used to improve teacher hiring.^v In a refreshing departure from much research on education reform, we find that DCPS has had considerable success in identifying and hiring effective teachers. Specifically, we find that several relatively inexpensive approaches to applicant screening implemented by the district strongly predict teacher performance in the classroom. Indeed, we conclude that DCPS principals should utilize the available measures of applicant quality even more than they currently do. More broadly, the experience of DCPS highlights the potential benefits that other districts might realize by developing a comprehensive, research-driven and thoughtful approach to teacher hiring.

Teacher selection in the DCPS

In 2009, DCPS created a centralized application process to streamline hiring by screening out less desirable applicants and giving principals a list of “recommended” candidates. On the basis of similar work we had done with New York City public school teachers,^{vi} DCPS officials asked us to help them to develop and evaluate various applicant assessments. The goal was to find measures that were both feasible to administer and able to predict which applicants would be most successful as classroom teachers. Over the next several years, the district developed and refined a variety of measures. During this period,

roughly half of new hires came through this centralized “TeachDC” selection process. Remaining new hires entered through alternative certification routes such as Teach for America, the DC Teaching Fellows, or were hired directly by a principal.

TeachDC has an online application system that collects background information such as applicants’ education history, employment experience, and eligibility for licensure.^{vii} Following collection of this preliminary information, applicants are asked to complete up to three additional in-depth assessments.^{viii} First, applicants take a subject-specific written assessment to evaluate their knowledge of subject matter content and associated instructional practices. District personnel score these written assessments and applicants who pass are then invited for a 30-minute structured interview covering areas such as their track record of success and how they respond to challenges. Interviews are also scored, and passing candidates are invited to a final “audition” stage where DCPS personnel watch them teach a lesson of approximately 30 minutes. District officials evaluate these lessons with the same classroom observation rubric used to evaluate current DCPS teachers.

This process is quite selective, with only 13-14 percent of applicants hired each year. Even among the set of applicants who make it through the entire process and are placed on the recommended list, only 50 percent are hired. While some teachers declined positions they were offered, the ratio of applicants to positions in DCPS means that the district can be quite choosy (though this does vary by content area).

Do application measures predict being hired in the district?

We analyzed which applicants were hired among over 7,000 individuals who applied through TeachDC in the years 2011-2013. We examined background characteristics (e.g. prior experience) and composite scores for applicant performance on the content knowledge assessment, the interview and the audition.^{ix}

Perhaps surprisingly, we found that applicants with better academic credentials were not more likely to be hired than other applicants. For example, neither an applicant’s SAT/ACT score nor her undergraduate GPA was associated with the likelihood of being hired. Importantly, this was not simply a function of better qualified applicants passing up DCPS to take a job in a more affluent district. Teachers who attended more

selective colleges were more likely to be hired, but this relationship was extremely small—i.e., moving one step up on Barron’s rank of college selectivity was associated with a mere 1 percentage point increase in the likelihood of being hired. Moreover, if we focus only on those candidates who reached the recommended pool, we find that applicants with higher GPAs and test scores were less likely to be offered a position. What can explain these results? It seems unlikely that principals lack good information on these characteristics since all candidates are required to submit a resume in addition to providing this information on the application itself. One possibility is that principals simply do not place much weight on a candidate’s academic performance because they view other personal qualities as more important in determining the effectiveness of a teacher. Indeed, our results here are consistent with some prior work finding that principals do not highly value traditional measures of academic success.^x

The pattern is much different when we examine applicants’ scores on the content knowledge assessments, the interview and the teaching audition. Higher scores on these measures *strongly* predict being hired.^{xi} Applicants who scored at the 83rd percentile of the content knowledge exam are roughly 40 percent more likely to be hired as applicants who scored at the 50th percentile. Applicants who scored at the 83rd percentile of the audition are *more than twice* as likely to be hired as applicants who scored at the 50th percentile. Together, these results suggest that principals value performance-based assessments more than academic credentials.

Do application measures predict job performance?

We measure teacher performance using the district’s teacher evaluation system (IMPACT), which scores every teacher in the district every year based on a number of measures, including classroom observation, principal assessment, and student achievement. This is a high stakes system. Teachers receive a single IMPACT score, and then are assigned one of five ratings based on this score: ineffective, minimally effective, developing, effective, and highly effective. Teachers in the “ineffective” category are immediately dismissed. Teachers are also dismissed if they fall in the “minimally effective” (or below) category for two consecutive years. On the other hand, teachers scoring in the “highly effective” category receive a one-time bonus of as much as \$25,000. If a teacher is rated highly effective for two consecutive years, she can

receive a substantial permanent increase in salary.

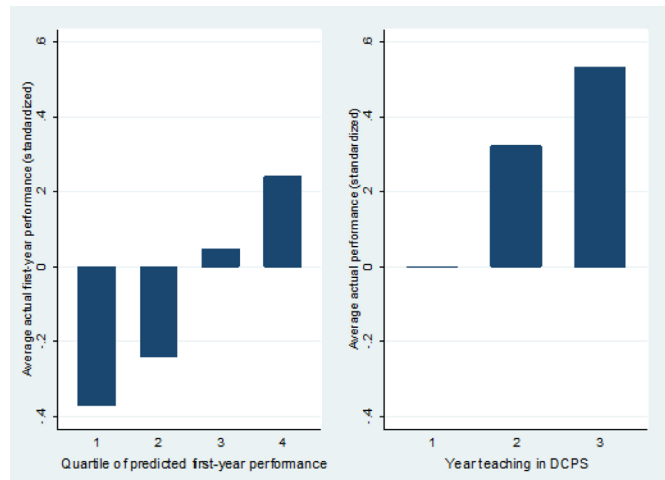
We found that a number of the measures collected as part of the application process were positively related to performance in the classroom.^{xii} Appendix Table A presents regression coefficients for the complete set of selection measures. Below I summarize some of the key findings.

Candidates with higher GPAs and those from more selective colleges performed systematically better in the classroom than otherwise identical candidates, even after accounting for the candidates' other application scores. This suggests that principals might want to place more weight on these characteristics than they have in the past.

In addition, we find that the content knowledge, interview, and audition scores are *all* strong, positive predictors of teacher performance. For example, the difference in performance between an applicant who scored at the 50th and 83rd percentile on the interview is equivalent to the difference in effectiveness between a first- and second-year teacher.^{xiii}

But what do these numbers mean for the overall quality of teachers hired? To illustrate the combined power of the different background characteristics and screening tests, we calculated a summary measure of applicant quality. Specifically, for each applicant, we calculate their predicted effectiveness using *all* of the information collected during the selection process. Figure 1 displays the average on-the-job performance for applicants, separately by their composite applicant score. In each case, performance is measured relative to the average first-year teacher. Candidates in the bottom quartile of applicant quality end up with teaching performance nearly 0.4 standard deviations below average. On the other end, candidates in the top quartile performed more than 0.2 standard deviations above average. To provide a point of comparison, we show the average performance of second- and third-year teachers in the district, again compared with the average first-year teacher. Second-year teachers performed about 0.3 standard deviations better than first-year teachers on average. Third-year teachers performed roughly 0.55 standard deviations better than first-year teachers. So, the difference between applicants in the top and bottom quartile is roughly equivalent to the benefit of two full years of experience.^{xiv}

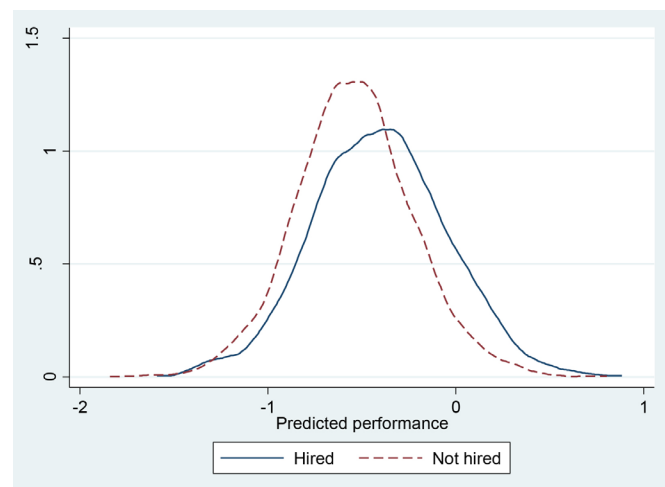
Figure 1. Classroom performance by teacher groups



Room to continue improving the hiring process

Our analysis suggests that the current teacher selection process in DCPS is helpful in differentiating between candidates that are likely to be more or less effective teachers. At the same time, we find that there still exists considerable scope for improving teacher quality in DCPS through the selection process. To illustrate this, we plot the distributions of predicted first-year performance separately for applicants hired and not hired in Figure 2. Teachers who are hired have significantly higher predicted effectiveness than those applicants who are not hired. Still, there is substantial overlap in the distributions. That is, there are many applicants who are *not* hired but whose predicted performance exceeds the average of those hired.

Figure 2. Predicted performance for applicants hired and not hired



One can also see this by examining the proportion hired as a function of the predicted quality score. Roughly 10 percent of applicants in the bottom third of predicted performance were hired. The likelihood of being hired does increase among those with higher predicted quality, but only 30 percent of the very best applicants (the top 5 percent) ended up working in DCPS.

This is due in part to the fact that once candidates reached the recommended pool, the application scores were much less important in determining who was offered a position. For example, recommended applicants with high content knowledge scores are no more likely to be offered a position than recommended applicants with low content knowledge scores. These results suggest that while high scores on these measures did help candidates reach the recommended list, principals did not rely on them to further distinguish between recommended candidates.

Our findings serve as a reminder that providing a list of candidates to principals may not be sufficient to guarantee top applicants are hired. Districts will need to pay attention to how principals access and how well they utilize this information. Until 2013, for example, there was no easy way for principals to access all of the selection data from the online system. To its credit, the district recognized this problem and has taken steps to help make the screening results more easily available to principals.^{xv}

Of course, it is important to consider the costs as well as the benefits of an improved teacher selection process. A thorough teacher selection system does entail some additional costs, but the DCPS experience suggests that such costs are quite small relative to the likely benefits. The primary additional cost of the DCPS system is the labor of administrators who conduct and score the interviews and teaching auditions (about one hour each for interviews and auditions) and staff to supervise and manage the process. We estimate that the total cost of the additional screening is no more than \$500 per hire.^{xvi} Compared with the cost of removing low-performing teachers, with large urban districts often reporting the cost of the dismissal process to be well above \$100,000,^{xvii} this expense is quite small. Moreover, relative to the anticipated long-run benefits to future students of hiring more effective teachers, this is undoubtedly a worthwhile investment.^{xviii}

The predictive power of the DCPS teacher selection system is encouraging, particularly in light of the many expensive and/or unsuccessful ways to improve the teacher workforce that have been tried to date. To be sure, selection is not a panacea. It is likely to be less beneficial in smaller markets, hard-to-staff fields or other contexts with fewer qualified applicants. And those doing the actual hiring may not fully appreciate or utilize the information collected as part of the selection process. Still, districts stand to gain considerably by investing in systems to identify the best possible applicants.

Appendix Table A. Relationship between application measures and teacher job performance

	Characteristics separately (1)	Characteristics simultaneously (2)
Years prior experience		
1 to 2	0.100 (0.090)	0.098 (0.083)
3 to 5	0.101 (0.099)	0.156+ (0.088)
6 to 10	0.034 (0.092)	0.064 (0.087)
11 or more	-0.214+ (0.117)	-0.109 (0.109)
Undergrad GPA (std)	0.243** (0.037)	0.185** (0.035)
SAT math+verbal (std)	0.155** (0.039)	0.019 (0.038)
Barron's Rank (linear 0-5)	0.149** (0.029)	0.109** (0.027)
Master's degree or higher	0.215** (0.064)	0.112+ (0.062)
Location of undergrad or grad school		
DC	0.020 (0.094)	-0.020 (0.087)
Maryland or Virginia	-0.085 (0.074)	-0.075 (0.069)
PCK written test (std)	0.260** (0.056)	0.186** (0.053)
Interview (std)	0.298** (0.055)	0.283** (0.051)
Audition (std)	0.149* (0.066)	0.104 (0.065)

Note: Estimates from least squares regressions with 1,581 teacher-by-year observations, and 917 unique teachers. The dependent variable is job performance measured by the first predicted factor from a factor analysis of IMPACT evaluation component scores, standardized. In column 1, each group of coefficients separated by a solid line are estimates from a separate regression. In column 2, estimates are from a single regression. Each specification includes recommended pool x year FE, year-by-subject-taught fixed effects, and indicators for second year in the district and third year in the district. When a covariate is missing for an observation, we set the value to zero and include an indicator = 1 for anyone missing that covariate. Clustered (teacher) standard errors in parentheses. + indicates $p < 0.10$, * 0.05, and ** 0.01

-
- ⁱ Dee, T., and Wyckoff, J. (2013). "Incentives, Selection, and Teacher Performance: Evidence from IMPACT." National Bureau of Economic Research, Working Paper No. 19529. Adnot, M., Dee, T., Katz, V. and Wyckoff, J. (2016). "Teacher Turnover, Teacher Quality, and Student Achievement in DCPS." National Bureau of Economic Research, Working Paper No. 21922.
- ⁱⁱ For the area of early elementary reading, for example, see: Slavin, R. E., Lake, C., Chambers, B., Cheung, A. and Davis, S. (2009). "Effective Reading Programs for the Elementary Grades: A Best-Evidence Synthesis." *Review of Educational Research*, 79(4): 1391-1466.
- ⁱⁱⁱ Boyd, D., Grossman, P., Lankford, H., Loeb, S. and Wyckoff, J. (2009). "Teacher Preparation and Student Achievement." *Educational Evaluation and Policy Analysis*, 31: 416-440.
- ^{iv} Shortages are more common in certain fields (e.g., special education, bilingual education, and some math and sciences) and in certain locations (e.g., rural areas).
- ^v For a more complete report on this study, see Jacob, B.J., Rockoff, J.E., Taylor, E.S., Lindy, B. and Rosen, R. (2016). "Teacher Applicant Hiring and Teacher Performance: Evidence from DC Public Schools." National Bureau of Economic Research, Working Paper No. 22054. As noted in the text, my colleagues and I started this project as unpaid consultants to DCPS, but the findings in this study do not necessarily reflect the position of DCPS. Funding for this research was provided by the Smith Richardson Foundation.
- ^{vi} Rockoff, J.E., Jacob, B.J., Kane T.J., Staiger D.O. (2011). "Can You Recognize an Effective Teacher When You Recruit One?" *Education Finance and Policy*. 6(1): 43-74.
- ^{vii} Applicants who don't already hold a DC license and whose credentials make them ineligible to obtain one prior to the start of the school year are not allowed to proceed further.
- ^{viii} The content of the application process has changed somewhat over time, though the three key assessments—subject-matter assessment, interview and audition—have always existed.
- ^{ix} Interestingly, while the correlations among application scores themselves are positive, they are all fairly low in magnitude, suggesting that each facet of the application captures distinct information about applicants. Of course, low correlations also may indicate a considerable amount of noise in each score.
- ^x See, for example, Ballou 1996 and Hinrichs 2014. Ballou, D. (1996). "Do Public Schools Hire the Best Applicants?" *Quarterly Journal of Economics*, 111(1): 97-133. Hinrichs, P. (2014). "What Kind of Teachers Are Schools Looking For? Evidence from a Randomized Field Experiment." Federal Reserve Bank of Cleveland, Working Paper 14-36.
- ^{xi} A one standard deviation increase in the content knowledge, interview and audition scores are associated with an increase in the likelihood of being hired by 6, 10.8 and 15.8 percentage points respectively (all relative to a baseline mean of 13 percent).
- ^{xii} Our results are robust when we account for selection into the district based on hiring.
- ^{xiii} None of the application measures described above had a significant relationship to whether (or when) a teacher left the district.
- ^{xiv} However, it is likely that the district could improve this process even more. For example, applicants in 2011 completed a 50-item assessment from the Haberman Star Teacher Pre-Screener, a commercial teacher applicant screening instrument that DCPS was piloting that year. Used by a number of large urban school districts throughout the U.S., the Haberman Pre-Screener is intended to provide school officials with guidance on how effective a particular candidate is likely to be in an urban classroom. Consistent with prior work in New York City, we found a strong positive relationship between teacher scores on the Haberman test and teacher performance in the classroom. The district discontinued the use of the Haberman after the pilot year, our results suggest that the district might consider reinstating the Haberman or other similar screens.
- ^{xv} For example, the district has created an online platform that allows school leaders to easily access candidate resumes as well as information on their performance in the selection process.
- ^{xvi} The district budgets roughly \$63,000 per year for interviews and auditions, and only \$7,500 for the prior applicant screening. There is an additional budget for staff to manage the recruitment and screening process, though much of that cost would be required even without the interviews and auditions.
- ^{xvii} Chait, R. (2010). "Removing Chronically Ineffective Teachers: Barriers and Opportunities." Center for American Progress.
- ^{xviii} Chetty, R, Friedman, J.N. and Rockoff, J.E. (2014). "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *American Economic Review*, 104(9): 2633-79.