Hiring Effective Teachers

Chantal L. McMahon
St. Catherine University

Hiring Effective Teachers

March 2014
Draft for Undergraduate Research Symposium
Important: Do not cite or circulate without author permission

Chantal McMahon
St. Catherine University
Undergraduate student collaborator

Dr. Kristine West (corresponding author)
St. Catherine University
Department of Economics

Dr. Aaron Sojourner
University of Minnesota
Department of Work and Organizations

Dr. Elton Mykerezi
University of Minnesota
Department of Applied Economics

Christopher Moore
Evaluation and Testing Specialist

Maggie Sullivan
Director of Strategic Planning

I. Introduction

This paper is a result of a new collaborative research effort between a large urban school district in the Midwest and a research team of economists from the University of Minnesota and St. Catherine University. The research uses district-defined teacher effectiveness measures to explore various human resource questions within the theoretical framework of labor and personnel economics. We define teacher effectiveness as a four-part measure in keeping with the district's current practice. Specifically, we use value-added measures in math and reading, teacher evaluations using standardized ratings of instructional effectiveness observed in classrooms and student surveys of teacher practice.

We plan to investigate a set of human resource questions related to recruiting and selecting the most effective teachers. We will consider whether the following factors predict teacher effectiveness: date of offer, date of hire, preparation program that led to initial licensure (e.g., traditional higher education institutions, Teach for America), student-teaching experience in the district, undergraduate or graduate grade point average, and other educational factors (e.g., graduate degree in field other than teaching). This analysis will use controls for teacher demographics (e.g. years of experience, education, gender, age), educational context (e.g. grade, year, school, subject) and student demographics (e.g. race/ethnicity, gender, English-language learner, free/reduced lunch status, age).

At this point in the project, we have one year of value added measures, one year of teacher evaluations and one year of student surveys along with data on date of hire, student-teaching experience (for new teachers only), grade-level, job title and student demographics. In this initial draft we focus on two questions: (1) Are teachers who have student taught in the district more effective than their peers? and (2) Does the month of hire predict teacher efficacy?

This is the first use of this district's data to investigate these important human resource questions. The results will be of interest to other districts as they develop their own measures of teacher effectiveness and think about how to use these data to better their hiring practices. The results will be of interest to the policy makers and the academic community because they will add to our knowledge of the predictors and correlates of various measures of effectiveness.

II. Previous literature

Teacher effectiveness has a significant impact on a student's success in school and in the labor market (Henry et al 2012). The value of a good teacher is hard to overstate. Recent estimates suggest that a teacher one standard deviation above the mean effectiveness generates over $400,000 in net present value (Hanushek 2011; Chetty et al 2011). Unfortunately, researchers have yet to identify measures visible at the time of hire that predict teacher effectiveness (Clotfelter, Ladd and Vigdor 2007; Hanushek and Rivkin 2012; Harris and Sass 2011). The variance between teacher effectiveness within schools is large, which suggests that hiring practices are not successfully differentiating between effective and ineffective teachers.

Absent clear signals of efficacy at the time of hire, schools must hire based on a noisy signal of effectiveness and then observe the teacher during a probationary period. In theory, the system is set up such that if during the probationary period the teacher proves to be effective, she is granted tenure and if she is ineffective, she is dismissed. In practice, almost all teachers are granted tenure. Research into the optimal number of years for the probationary period and the optimal number of teachers granted tenure suggests that the current practice is far from ideal. For instance, Staiger and Rockoff (2010), show that absent useful information about effectiveness at the time of hire, dismissal rates may need to be in excess of 80%. They also clearly show that as the reliability of a pre-hire signal increases, the proportion of teachers dismissed should fall.

In the paper, we argue that there is a very reliable pre-hire signal that is largely overlooked. There has been little investigation into whether student teaching is a reliable signal of teacher efficacy. Boyd et al (2009) include two student teaching variables in their analysis of teacher preparation. Specifically, they test whether student teaching and the degree to which student teaching matches the current job title predict teacher value added. They find that neither having student taught nor having student taught in a grade/subject similar to your own is strongly predictive of teacher value added. We ask a slightly different question: whether student teaching in your current district is predictive of teacher effectiveness?

Student teaching is usually thought of as the final experience before entering the labor market but it also the first experience to observe the teacher on the job. Generally, evaluation of student teachers is the purview of institutions of higher education. School districts are involved only in so much as they provide a setting for this capstone experience.[1] School districts need to think of student teaching as an extended -- and potentially very informative -- job interview. Student teachers often work in a district for an entire semester and they work closely with the district's teachers and students. There is ample opportunity to gather information about the teacher that is not available on a resume or made visible during a traditional interview. Anecdotally, we observe that districts do not systematically make use of this information. However, even ad-hoc sharing of this information should make it such that the candidates hired from the pool of student teachers are of higher quality than the candidates hired from the general pool of applicants.[2]

In addition to having a more complete picture of the applicants abilities, districts might find that if an applicant applies to the same district where she student taught this indicates a strong commitment to that district and/or reveal that there is a particularly good match between the teacher and the district. Jackson (2013) finds that teacher effectiveness increases after a transfer to a different school, indicating the effect of match-quality. Those with high match-quality are more likely to stay with the school than those with low match-quality. Therefore we expect that teachers hired from the pool of student teachers will be more likely to remain with the district than teachers hired from the general pool of applicants.

Turning to month of hire, although they often have more than enough applicants to fill vacancies, urban districts tend to hire late forcing many of the applicants to withdraw and accept other offers from schools with a quicker response time (Levin and Quinn 2003). We do not know of any research that links

[1] For example, a study by the National Council on Teacher Quality, Greenberg, Pomerance and Walsh (2011), that looks at student teaching, is based on the question of whether teacher preparation programs are doing an adequate job structuring and assessing this experience.
[2] In this way our work is in conversation with a larger literature on employer learning, i.e. Altonji and Pierret (2001).

date of hire to teacher efficacy as measured by value-added, evaluations and/or student surveys. This paper will fill that void.

III. Data & Methodology

The data are administrative data compiled from various departments within the district. There are over 2,500 classroom teachers currently employed in the district (2013-2014 school-year).[3] We have date of hire, full/part time, job title and highest degree for all of them. Table 1 column 1 summarizes these variables. Ninety two percent of the teachers work full time. The most frequent job title is "teacher, elementary" followed by "teacher, special education." The average teacher has been in the district for 12.5 years and is on the 16th step of the salary schedule.[4] The average teacher is between the 9th and 10th lane on the salary schedule which corresponds to a Masters degree and 15 additional college credits earned after the Masters. Figure 1 shows the full distribution of experience with the district. There are many teachers who have been hired in the last four years. There is a dip in the distribution that suggests that in the early 2000s there was either a dip in hiring, or teachers who were hired then left the district at a higher than average rate.

We know the grade-level(s) they are currently teaching as well as the demographic makeup of their students for 1,483 teachers. Table 2 summarizes these variables. Over 2/3 of the districts' students are eligible for free or reduced lunch. Approximately one third of the district's students are African American and approximately 1/3 are white. Just over ¼ of the district's students are English language learners (ELL). Figure 2 shows the distribution of teachers by grade. Some teachers are multi-counted in this figure because they teach multiple grades. We see that we are missing a number of 1st and 2nd grade teachers but otherwise have a reasonably uniform distribution across the grades with a slight uptick in high school. Linking student information to teachers for the current school year was accomplished via the online student survey portal. Few 1st and 2nd grade teachers are included because their students took the survey predominately on paper. First and second graders will eventually be linked to teachers through the grading system and through a roster verification process prior to estimating the value added by individual teachers.

---

[3] We have data for over 3,000 individuals but we exclude job titles such as occupational therapist and school nurse. Our definition of classroom teachers includes teachers who work with small groups of students such as special education teachers.

[4] In the results section we use step rather than years in the district to measure experience. Step is a better control for overall teaching experience because teachers who transfer in to the district are often granted credit for their experience in other districts by being placed at a higher step on the salary schedule.

We have value-added measures from the previous school-year in math and reading for 897 and 948 teachers respectively. We have teacher observation scores from the previous school-year for 2,444 teachers and student survey scores from the current school year for 2,066 teachers. Table 3 summarizes these variables and Table 4 displays the correlation between them for the 583 teachers for which we have all four measures. The value-added scores are calculated by the district's research and evaluation staff and the Value Added Research Center (VARC). For this research project, we only observe the resulting score for each teacher that has sufficient students in each tested subject-grade. The teacher evaluation scores are the result of four observations by administrators and teacher-leaders all of whom have been trained using a standardized observation rubric referred to as the "standards of effective instruction (SOEI)." This rubric is specific to the district but borrows heavily from the Charlotte Danielson (1996) evaluation framework. The student surveys were initially based on the publicly available version of the Tripod Seven C's survey of teacher practice (Kane and Staiger 2012). Questions were revised and added in collaboration with the district's teacher evaluation advisory group. We refer to SOEI observations and student surveys as "subjective evaluations." This is consistent with the use of the term in labor and personnel economics literature. However, it is worth emphasizing that these measures have been carefully developed so, while they are "rater mediated," they are designed with consistency and objectivity in mind.

In all cases, teacher performance is measured with error. For each measure, we quantify the proportion of variance in observed scores attributable to teachers (i.e., reliability). Reliability is then used to estimate each teacher's "true" score and standard error of measurement for research and reporting purposes. The reliability of observation scores is about 0.65. The reliability of student survey scores is 0.81 for K-2 teachers, 0.82 for grades 3-5, and 0.90 for grades 6-12. The reliability of value-added scores ranges from 0.44 (grade 7 reading) to 0.9 (grade 6 math). Appendix A contains more information on reliability.

We know of fifteen teachers hired in the last two years who were previously student teachers in the district. Tables 1, 2 and 3 compare these teachers to other new teachers and to all other teachers in the district. Table 3 shows that the teachers who student taught in the district score slightly higher on the student surveys than other new teachers, although the difference is not statistically significant.[5] This suggests that student teachers may be of higher quality than other recent hires. Of course, these results are merely suggestive and in the next section we control for a variety of covariates. For instance we also see that student teachers differ from other new teachers in that student teachers are disproportionally special

[5] Because these are new teachers and value added scores and teacher evaluations are only available for previous years, we focus on student surveys.

education teachers and are slightly more likely to be working full time for the district. On average, student teachers are on the 3rd step on the salary schedule and, on average, other new teachers are on the fifth step. This is because student teachers are, by definition, new graduates while new teachers may be new graduates or they may be transfers from another district who are credited for that experience by being placed at a higher step.

Table 2 shows that all new teachers, regardless of whether they student taught in the district, have a larger share of African American students and more students who are free and reduced lunch eligible than their peers who have been with the district at least one year. This is consistent with literature that shows that at-risk students face more teacher turnover and are more likely to have an inexperienced teacher than their more advantaged peers (Feng 2010).

Turning to date of hire, we see that majority of teachers have a start date in August. Although the majority of teachers start in August, fully 38% begin working after the school year has already begun. Figure 3 shows the distribution of teacher start dates. Unfortunately, at this point we do not know the date the offer was extended but rather the date that the contract officially starts. However, intuitively there are no advantages for a teacher to begin mid-semester. This implies that those who begin late have received a late offer from the district and may not be as effective as teachers who receive offers immediately. The summary statistics show that student teachers are more likely to be hired in August than other new teachers. This suggests that student teachers receive offers earlier than non-student teachers, indicating that they may be more desirable applicants. Tables 1, 2 and 3 separate out teachers who start in August and they do not appear to differ from the full set of teachers along any observable dimension. They score slightly higher on the measures of teacher effectiveness but the differences are not statistically significant.

For the results in the next section, we use each of the measures of effectiveness -- student surveys, teacher evaluations, value-added math and value-added reading – as dependent variables in simple OLS regressions. Specifically, we estimate:

$$Y_i = \alpha + \gamma S_i + \delta A_i + \beta X_i + \theta T_i + \varepsilon \tag{1}$$

Where Y_i is one of the four measures of effectiveness for teacher i, S_i is an indicator that equals one if the teacher was a student teacher in the district, A_i is an indicator that equals one if the teacher was hired to start in August and X_i is a vector of controls that include indicators for being new to the district, being full time, holding multiple positions, job title and grade level along with continuous variables that measure the share of the teacher's students who are ELL, free/reduced lunch eligible, highly mobile, black, native, Asian, white, and female. Lastly, T_i is a vector of continuous variables that measure the teacher's current step and lane placement. Step is controlled for as linear, squared, cubed and quartic to allow for a flexible functional form.

When the dependent variable measures student surveys or teacher evaluations, we use a standardized variable that has been transformed to have a mean of zero and standard deviation of one. This eases interpretation of the coefficients in the regression. Value added scores were not standardized because, by design, they have a mean of zero and a standard deviation.[6]

The main coefficients of interest are γ, which shows whether teachers who student taught in the district outperform their peers and δ, which shows whether teachers who start in August outperform their peers. In each case the magnitude of the coefficient measures how many standard deviations these teachers are above (below) their peers. The estimation is cross-sectional in nature and thus cannot speak to trends over time. Also, the simple OLS estimation strategy does not provide for a convincing counterfactual so the results should be interpreted as mere associations rather than causal.

IV. Preliminary Results

Table 5 presents results for equation (1) where the dependent variable is a subjective measure of teacher effectiveness, namely a formal evaluation or the student survey. Table 6 presents results for equation (1) where the dependent variable is a value-added score, either in math or in reading. We present results with and without controls for grade level and student demographics because we do not have that information for all teachers and thus including these controls reduces the sample size.

The only outcome for which we can assess teachers who were student teachers in the district is student surveys. In column (d) of Table 5, we see that teachers who student taught in the district score, on average, 0.346 standard deviations higher than their peers. The magnitude is smaller without controls for grade level and student demographics and in both specifications the result is not statistically significant. The small number of student teachers likely explains the imprecision of the estimate. It is interesting that the magnitude of the coefficient is roughly opposite to the magnitude of the coefficient on being new to the district, -0.358. This is consistent with the story that student teachers come in to the district with approximately one year's worth of district specific human capital and thus "hit the ground running" in a way that their peers who student taught elsewhere cannot.

[6] To be exact, the district calculates value added scores to have a mean of zero and then adds +3 to all scores so that no teachers are shown a "negative" value added. District officials were concerned that a negative number would be interpreted as decreasing student achievement rather than scoring below average.

It will be interesting to see in future iterations of this paper, whether the result holds up to a larger sample of student teachers and whether the result is similar across other measures of teacher effectiveness. Interestingly, teachers who are new to the district do not seem to score much lower on formal evaluations, or value-added measures relative to their peers who have been in the district for at least one year.

Turning to start date, we find that teachers who start in August score higher on student surveys and formal evaluations than their peers who were hired to start in other months. The statistical significance on this result, however, is inconsistent across specifications. In the larger sample, columns (a) and (c), we find that teachers who start their careers with the district at the start of the school year, outperform their peers by about 0.07 to 0.08 of a standard deviation on subjective measures and 0.02 standard deviations on value-added measures. The value-added results, however, are statistically indistinguishable from zero in both samples/specifications presented here.

The coefficients on step and lane tell an interesting story. The relationship between step and effectiveness is consistent across all four measures. In every case, the linear term is positive, the squared term is negative, the cubed term is positive and the quartic term is negative. The coefficients on lane are very small and generally statistically insignificant. This supports findings of other researchers that have concluded that education (as measured by degrees and credits) is not strongly correlated with teacher efficacy.

The variables that measure the context in which the teacher teaches (i.e. her job title, full time status, grade level, and student demographics) are included primarily as controls. Some of these variables have been cut from Table 5 and Table 6 to save space. These are available upon request and are generally unremarkable. One exception… student surveys in lower elementary grades. The student demographics are included on Table 5 and Table 6. Here we see that student demographics seem to have the strongest association with value-added scores in reading. Having classes that are largely made up of black, native, Asian, or Hispanic students is associated with scoring a full standard deviation lower on this measure of effectiveness. This is not the case for value-added scores in math.

Also, in reading but not in math, large shares of ELL students or students who are eligible for free and reduced lunch are associated with higher teacher value-added. Interestingly, the coefficients on free and reduced lunch for the subjective measures, evaluations and student surveys, are negative. This suggests that teachers with large shares of low income students, subjective evaluations and value-added are potentially orthogonal. As noted above, reading value-added scores are less reliable than math value-added scores. Given that contextual factors explain variation in reading value-added scores but not in math suggests that the reading value-added model specification may be improved by including contextual factors. As it currently stands, the value-added models only control for individual, student-level characteristics and not peer effects.

V. Discussion

We expect that teachers who student taught in the district will be more effective than their peers. Student teaching may serve as a sorting mechanism or a year in which the teacher acquires district specific skills ahead of her peers who student taught elsewhere. Clearly teachers cannot student teach in every district that they apply for. Each district will observe a reliable signal for a subset of its potential hires. They should find that the applicants that they hire from this smaller pool are more effective than the applicants that they hire for the larger pool. We see some evidence that this is the case. We caution that we cannot make a strong claim because we have a very small sample of teachers who student taught in the district where they are currently working.

We expect that teachers hired to start in August are more effective than the applicants who are hired at other times. We find evidence that teachers hired to start in August score in the range of 7 to 10 percent of a standard deviation higher than their peers on subjective evaluations by both trained classroom observers and by students. We do not find much evidence that this is also the case for value-added measures of teacher effectiveness. Teachers hired to start in August score 1 to 8 percent of a standard deviation higher than their peers on these measures but these results are more statistically suspect.

Lastly, it is important to put our findings in perspective for policy applications. We do not mean to suggest that a district adopt a policy of hiring all the applicants who student taught in the district. It may be that hiring someone who student taught in the district is always better than hiring a teacher who student taught elsewhere either because they have acquired district specific human capital or because they are more attached to the district, but it is more likely that the applicants who student taught in the district who are passed over for a job, are passed over for good reason. If districts are making full use of the information obtained during student teaching, they should be hiring only the most effective student teachers. Our policy recommendation is only that districts make full use of the information obtained during student teaching. Likewise, hiring a teacher in August who would have otherwise started in some other month will probably not cause an increase in her classroom effectiveness in the long term. In the year hired, there is likely an advantage to starting at the beginning of the years, however, in subsequent years this advantage should disappear. The policy recommendation to hire earlier is based on the fact that hiring earlier provides better access to the full pool of applicants and thus allows a district to select the most promising from the group.

Next steps:

- Gather more student teaching data
- Add teacher demographic data
- Add teacher pre-hire data
- Look at teacher absences as an outcome
- Explore the other questions alluded to in the introduction

Bibliography

Altonji, J. and C. Pierret (2001). *Employer Learning and Statistical Discrimination.* The Quarterly Journal of Economics, 116(1), 313.

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

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2006). *How changes in entry requirements alter the teacher workforce and affect student achievement.* Education, 1(2), 176-216.

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2009) *Teacher Preparation and Student Achievement.* Educational Evaluation and Policy Analysis, 31(4), 416-440.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). *How does your kindergarten classroom affect your earnings?* Evidence from Project STAR. The Quarterly Journal of Economics, 126(4), 1593-1660.

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

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2010). *Teacher credentials and student achievement in high school a cross-subject analysis with student fixed effects.* Journal of Human Resources, 45(3), 655-681.

Cohen-Vogel, L. (2011). *"Staffing to the Test" Are Today's School Personnel Practices Evidence Based?.* Educational Evaluation and Policy Analysis, 33(4), 483-505.

Danielson, C. (1996). *Enhancing Professional Practice: A Framework for Teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.

Dee, T., & Wyckoff, J. (2013). *Incentives, selection, and teacher performance: Evidence from IMPACT* (No. w19529). National Bureau of Economic Research.

Dobbie, W. (2011). *Teacher characteristics and student achievement: Evidence from Teach For America.* Harvard University, July.

Feng, L. (2010). *Hire today, gone tomorrow: New teacher classroom assignments and teacher mobility.* Education, 5(3), 278-316.

Feng, L., & Sass, T. (2012). *Teacher quality and teacher mobility*. Andrew Young School of Policy Studies Research Paper Series, (12-08).

Goldhaber, D. D., & Hansen, M. (2008). *Assessing the Potential of Using Value-Added Estimates of Teacher Job Performance for Making Tenure Decisions.* Brief 3. National Center for Analysis of Longitudinal Data in Education Research.

Goldhaber, D. D., & Hansen, M. (2008). *Is it Just a Bad Class?: Assessing the Stability of Measured Teacher Performance*. Center on Reinventing Public Education.

Greenberg, J., L. Pomerance and K. Walsh (2011). *Student Teaching in the United States.* National Council on Teacher Quality report.

Hanushek, E. A. (2011). *The economic value of higher teacher quality.* Economics of Education Review, 30(3), 466-479.

Hanushek, E. A., Kain, J. F., O'Brien, D. M., & Rivkin, S. G. (2005). *The market for teacher quality* (No. w11154). National Bureau of Economic Research.

Hanushek, E. A., & Rivkin, S. G. (2007). *Pay, working conditions, and teacher quality*. The future of children, 17(1), 69-86.

Hanushek, E. A., & Rivkin, S. G. (2010). *Generalizations about using value-added measures of teacher quality*. The American Economic Review, 267-271.

Hanushek, E. A., & Rivkin, S. G. (2012). *The distribution of teacher quality and implications for policy.* Annu. Rev. Econ., 4(1), 131-157.

Harris, D. N., Rutledge, S. A., Ingle, W. K., & Thompson, C. C. (2010). *Mix and match: What principals really look for when hiring teachers.* Education, 5(2), 228-246.

Harris, D. N., & Sass, T. R. (2011). *Teacher training, teacher quality and student achievement.* Journal of public economics, 95(7), 798-812.

Henry, G. T., Fortner, C. K., & Bastian, K. C. (2012). *The effects of experience and attrition for novice high-school science and mathematics teachers*. Science, 335(6072), 1118-1121.

Herrmann, M. A., & Rockoff, J. E. (2010). *Worker absence and productivity: Evidence from teaching* (No. w16524). National Bureau of Economic Research.

Hinrichs, P. (2013) *What kind of teachers are schools looking for?* Working paper.

Huang, F., & Cappelli, P. (2010). *Applicant screening and performance-related outcomes*. The American Economic Review, 214-218.

Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Seattle, WA: Bill & Melinda Gates Foundation. Retrieved from http://www.metproject.org/downloads/MET_Gathering_Feedback_Research_Paper.pdf

Koedel, C., Parsons, E., Podgursky, M., & Ehlert, M. (2012). *Teacher Preparation Programs and Teacher Quality: Are There Real Differences Across Programs?*. National Center for Analysis for Longitudinal Data in Education Research, Working Paper, 79.

Jackson, C. K. (2012). *Do High-School Teachers Really Matter?*. National Bureau of Economic Research.

Jackson, C. K. (2013). *Match quality, worker productivity, and worker mobility: Direct evidence from teachers.* Review of Economics and Statistics.

Jackson, C. K., & Bruegmann, E. (2009*). Teaching students and teaching each other: The importance of peer learning for teachers*. American Economic Journal: Applied Economics, 1(4), 85-108.

Jacob, B. A. (2007*). The challenges of staffing urban schools with effective teachers.* The Future of Children, 17(1), 129-153.

Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2011). *Identifying effective classroom practices using student achievement data.* Journal of Human Resources, 46(3), 587-613.

Levin, J., & Quinn, M. (2003). *Missed Opportunities: How We Keep High-Quality Teachers out of Urban Classrooms.* The New Teacher Project.

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

Podgursky, M., Monroe, R., & Watson, D. (2004). *The academic quality of public school teachers: An analysis of entry and exit behavior*. Economics of Education Review, 23(5), 507-518.

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

Rockoff, J. E., & Speroni, C. (2010). *Subjective and objective evaluations of teacher effectiveness.* The American Economic Review, 261-266.

Rutledge, S. A., Harris, D. N., Thompson, C. T., & Ingle, W. K. (2008). *Certify, blink, hire: An examination of the process and tools of teacher screening and selection*. Leadership and Policy in Schools, 7(3), 237-263.

Staiger, D. O., & Rockoff, J. E. (2010). *Searching for effective teachers with imperfect information*. The Journal of Economic Perspectives, 24(3), 97-117.

Tables and Figures

Table 1: Employment Information

	All Teachers (N=2725)	New Teachers (N=356)	Student Teachers (N=15)	August Start Date (N=1687)
Job				
Full Time	0.92	0.88	0.93	0.93
Multiple Jobs	0.02	0.03	0.07	0.03
Job Title				
Elementary	0.30	0.29	0.20	0.29
Special Ed	0.21	0.23	0.47	0.20
ESL	0.06	0.06	0.13	0.06
Math	0.06	0.08	0.00	0.07
English	0.05	0.03	0.00	0.05
Social Studies	0.05	0.04	0.07	0.04
Science	0.04	0.04	0.07	0.05
Gym	0.03	0.03	0.00	0.03
Music	0.03	0.02	0.00	0.02
Language	0.03	0.04	0.07	0.03
Other	0.15	0.14	0.00	0.16
Degree				
Step	16.05 (9.65)	5.71 (5.88)	3.27 (3.35)	15.10 (9.41)
Years in District	12.52 (9.64)	0.9 (0.31)	0.93 (0.26)	11.54 (9.24)
Lane	9.50 (5.88)	4.92 (3.67)	4.47 (2.23)	9.29 (5.82)
Hire Month				
August	0.62	0.65	0.87	N/A
Other	0.38	0.35	0.13	N/A

Footnote: Mean (SD).

Table 2: Student Demographics

	All Teachers (N=1483)		New Teachers (N=169)		Student Teachers (N=7)		August Start Date (N=965)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Af.Amer.	0.36	0.25	0.45	0.29	0.44	0.30	0.35	0.25
White	0.30	0.28	0.20	0.24	0.22	0.32	0.30	0.29
Hispanic	0.21	0.24	0.22	0.25	0.22	0.23	0.21	0.25
Asian	0.09	0.15	0.07	0.11	0.03	0.06	0.10	0.16
Native	0.04	0.10	0.07	0.16	0.08	0.10	0.04	0.11
Mobile	0.05	0.09	0.07	0.11	0.17	0.19	0.06	0.09
Female	0.49	0.15	0.49	0.18	0.53	0.24	0.49	0.14
ELL	0.26	0.29	0.26	0.30	0.34	0.45	0.26	0.29
FRL	0.68	0.31	0.79	0.26	0.77	0.40	0.68	0.31

Table 3: Outcome Variables

		All Teachers	New Teachers	Student Teachers	August Start Date
Value-Added Math	Mean	2.98	2.93	N/A	3.01
	SD	0.82	0.91	N/A	0.84
	N	897	14	0	482
Value-Added Reading	Mean	3.01	3.10	N/A	3.03
	SD	0.69	0.64	N/A	0.68
	N	948	16	1	510
SOEI	Mean	2.92	2.82	N/A	2.94
	SD	0.22	0.14	N/A	0.21
	N	2444	31	1	1326
Survey	Mean	0.82	0.80	0.82	0.82
	SD	0.07	0.08	0.07	0.07
	N	2066	232	13	1296

Table 4: Correlations Between Measures of Teacher Effectiveness

N=583: Teachers with all four measures				
	Student Survey	**Formal Evaluation**	**Value-Added Math**	**Value-Added Reading**
Student Survey	1			
Formal Evaluation	0.0819	1		
Value-Added Math	0.0597	0.1326	1	
Value-Added Reading	0.0443	0.1167	0.5111	1

All positive but surprisingly low(?) If I do these for all teachers with the two relevant measures (rather than just the 583 with all measures), I find roughly the same correlations with the exception of student survey and formal evaluation which are more strongly correlated (0.1720).

Footnote… We conducted a factor analysis using the 583 teachers with all four measures. We plan to talk about this more at the conference. It seems that there are two latent dimensions of effectiveness that are picked up by these measures. One factor loads heavily on the two value-added measures but is positively correlated with all four measures. A second factor loads on the subjective measures – observations and student surveys – and is negatively correlated with value-added scores.

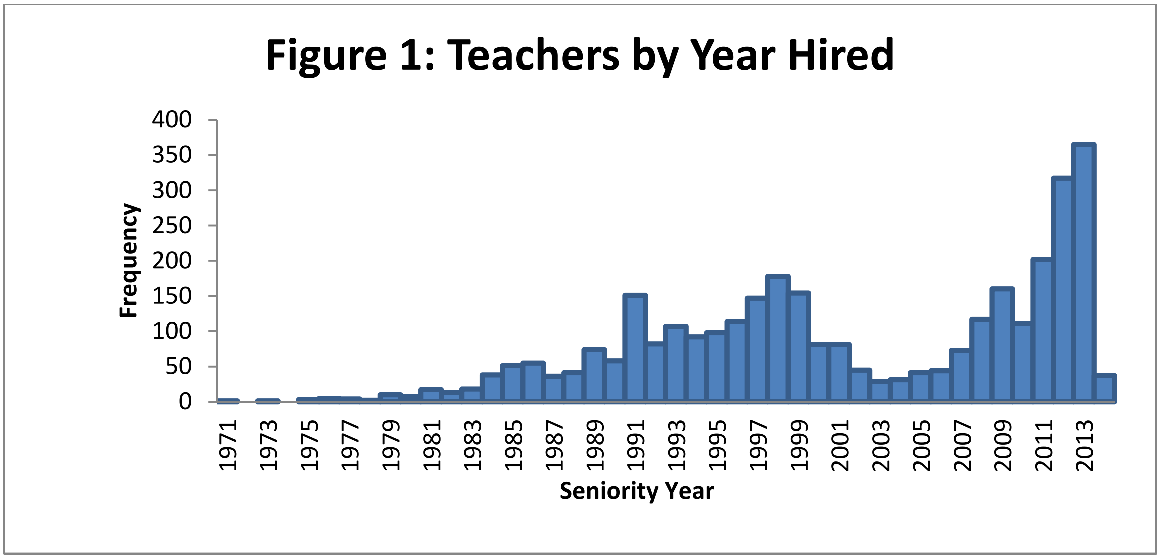


Figure 1: Teachers by Year Hired
Frequency
400
350
300
250
200
150
100
50
0
1971
1973
1975
1977
1979
1981
1983
1985
1987
1989
1991
1993
1995
1997
1999
2001
2003
2005
2007
2009
2011
2013
Seniority Year

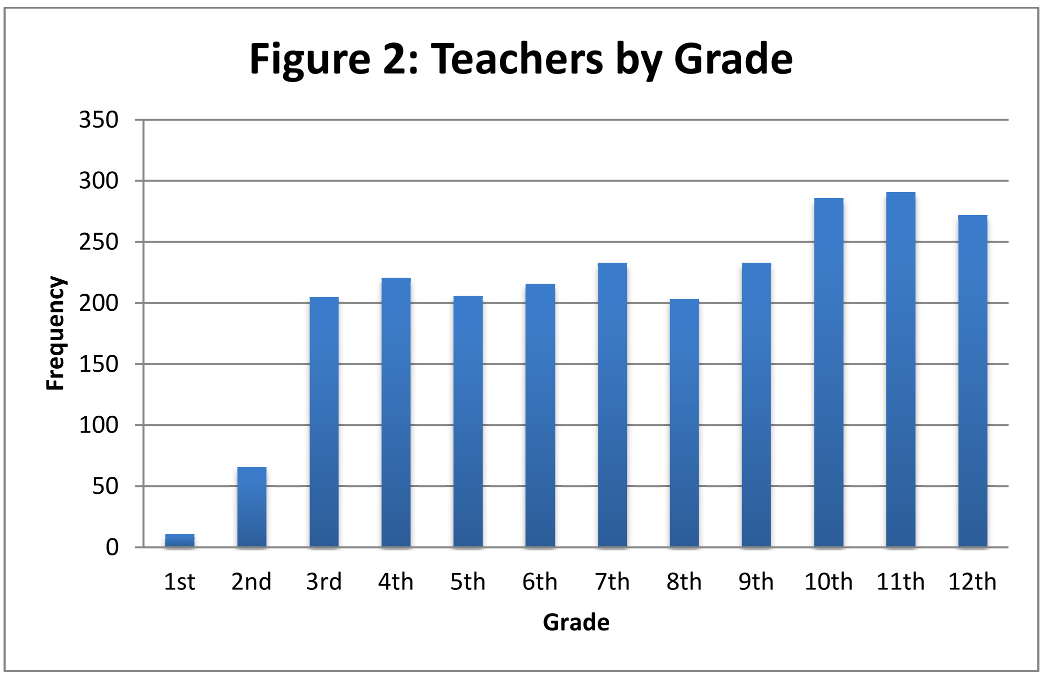


Figure 2: Teachers by Grade
Frequency
350
300
250
200
150
100
50
0
1st
2nd
3rd
4th
5th
6th
7th
8th
9th
10th
11th
12th
Grade

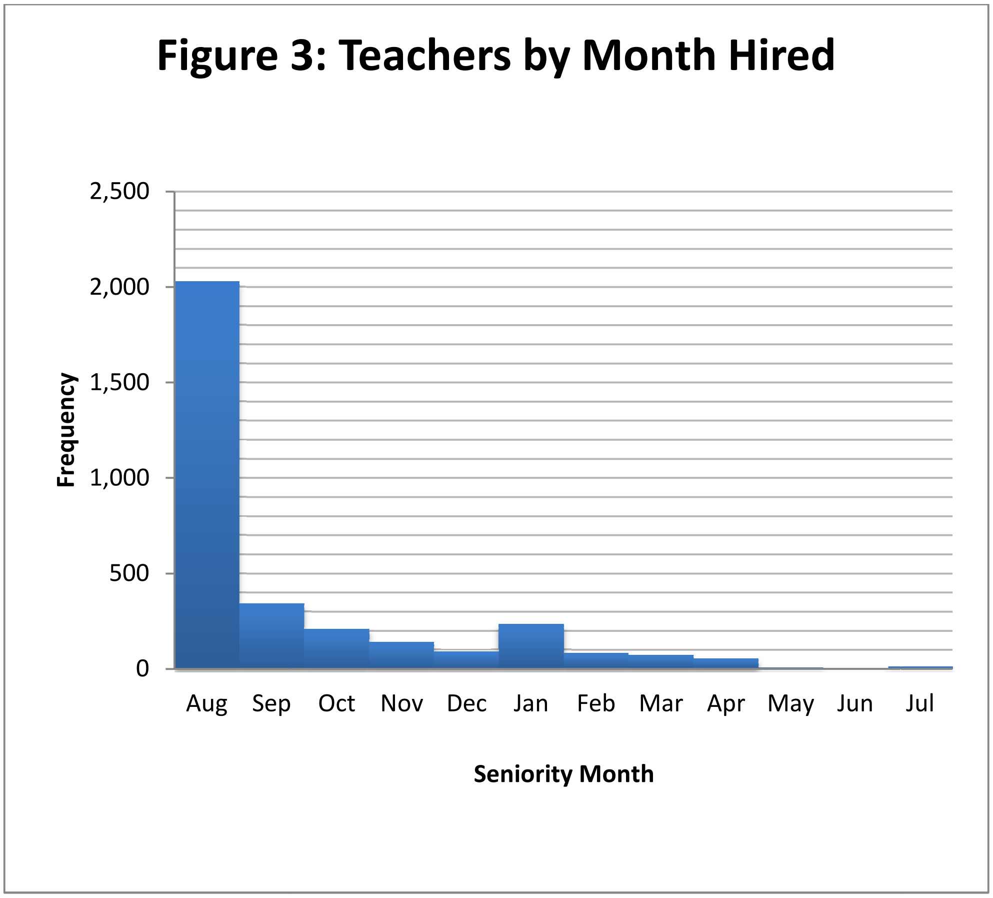


Figure 3: Teachers by Month Hired
Frequency
2,500
2,000
1,500
1,000
500
0
Aug
Sep
Oct
Nov
Dec
Jan
Feb
Mar
Apr
May
Jun
Jul
Seniority Month

Table 5: Subjective evaluation

	(a)	(b)	(c)	(d)
Dependent variable:	Evaluation		Student survey	
Started in August	0.0811**	0.0829	0.0719	0.103*
	(0.0405)	(0.0512)	(0.0474)	(0.0551)
Student taught in district	n/a	n/a	0.212	0.346
			(0.281)	(0.372)
New to district	-0.00794	-0.209	-0.328***	-0.358***
	(0.165)	(0.242)	(0.0836)	(0.0970)
Step	0.343***	0.302***	0.0751*	0.100**
	(0.0459)	(0.0564)	(0.0451)	(0.0510)
Step^2	-0.0273***	-0.0229***	-0.00899*	-0.0110**
	(0.00449)	(0.00565)	(0.00470)	(0.00531)
Step^3	0.000902***	0.000729***	0.000349*	0.000408*
	(0.000169)	(0.000216)	(0.000186)	(0.000210)
Step^4	-1.03e-05***	-8.11e-06***	-4.31e-06*	-4.88e-06*
	(2.14e-06)	(2.79e-06)	(2.46e-06)	(2.77e-06)
Lane	0.00742*	0.00449	0.00147	-0.00274
	(0.00440)	(0.00557)	(0.00527)	(0.00607)
Share ELL		0.0980		0.509***
		(0.153)		(0.158)
Share free/red lunch		-0.406*		-0.506*
		(0.244)		(0.263)
Share Afr.Amer.		-0.313		-0.0162
		(0.275)		(0.295)
Share Native		-0.638*		-0.486
		(0.362)		(0.364)
Share Asian		-0.156		-0.0128
		(0.286)		(0.307)
Share Hispanic		-0.264		0.149
		(0.293)		(0.312)
Share highly mobile		-0.434		0.292
		(0.343)		(0.354)
Share female		0.254		0.382**
		(0.187)		(0.193)
Constant	-2.035***	-0.941**	-0.285	-0.565
	(0.185)	(0.410)	(0.313)	(0.366)
Job information	yes	yes	yes	yes
Grade level	no	yes	no	yes
Observations	2,107	1,228	1,932	1,415
R-squared	0.118	0.208	0.065	0.167

Standard errors in parentheses

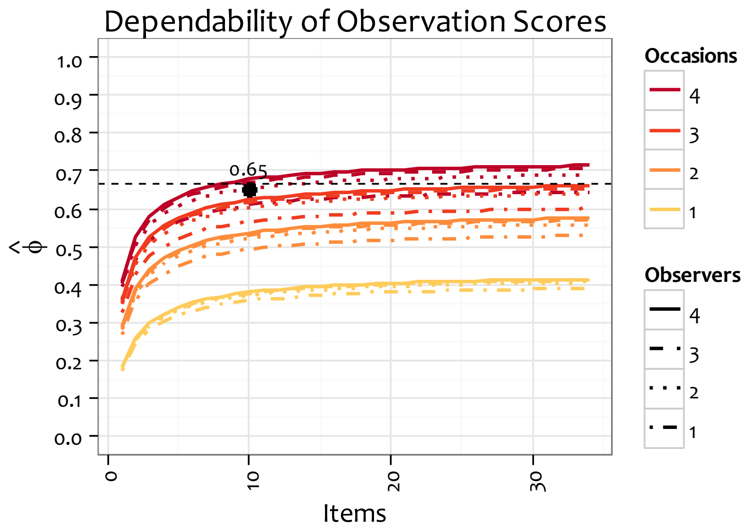
*** p<0.01, ** p<0.05, * p<0.1

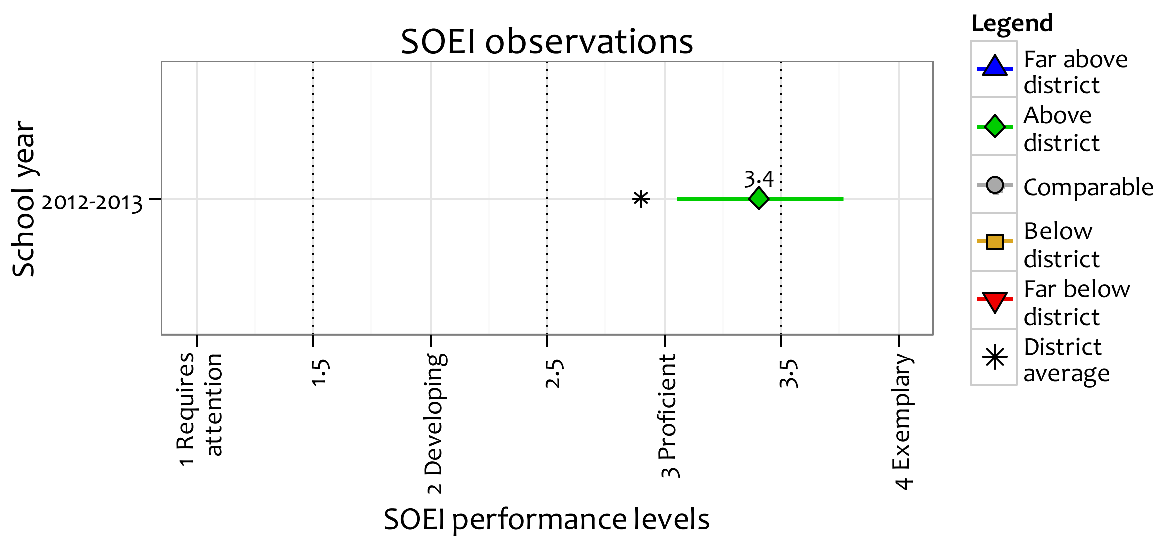
Table 6: Value added

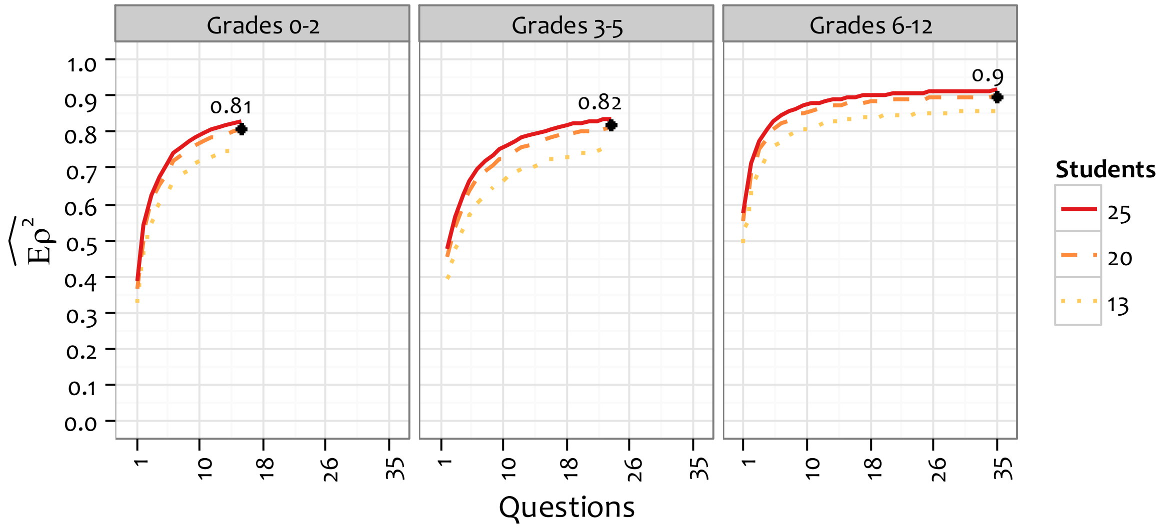
	(a)	(b)	(c)	(d)
Dependent variable:	Math		Reading	
Start in August	0.0230	0.00824	0.0239	0.0847
	(0.0616)	(0.0924)	(0.0507)	(0.0688)
New to district	0.0800	-0.147	0.190	0.508
	(0.226)	(0.487)	(0.181)	(0.335)
Step	0.125*	0.218**	0.120*	0.0729
	(0.0747)	(0.106)	(0.0625)	(0.0774)
Step^2	-0.0121*	-0.0198*	-0.0108*	-0.00381
	(0.00730)	(0.0104)	(0.00613)	(0.00761)
Step^3	0.000455*	0.000716*	0.000374	5.60e-05
	(0.000276)	(0.000393)	(0.000233)	(0.000288)
Step^4	-5.74e-06	-8.74e-06*	-4.39e-06	8.71e-08
	(3.54e-06)	(5.05e-06)	(2.99e-06)	(3.69e-06)
Lane	0.00275	-0.0101	0.00450	-0.00238
	(0.00679)	(0.00972)	(0.00586)	(0.00767)
Share ELL		0.237		0.461**
		(0.294)		(0.199)
Share free/red lunch		-0.0899		0.595*
		(0.448)		(0.341)
Share Afr.Amer.		0.299		-1.081***
		(0.402)		(0.384)
Share Native		0.000644		-0.944**
		(0.524)		(0.437)
Share Asian		0.263		-1.322***
		(0.482)		(0.414)
Share Hispanic		0.546		-1.177***
		(0.588)		(0.406)
Share highly mobile		0.662		-0.136
		(0.629)		(0.406)
Share female		0.307		0.246
		(0.452)		(0.319)
Constant	2.306***	1.466*	2.484***	2.673***
	(0.358)	(0.844)	(0.286)	(0.471)
Job information	yes	yes	yes	yes
Grade level	no	yes	no	yes
Observations	744	377	800	420
R-squared	0.018	0.051	0.012	0.085

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix A: Psychometrics







Grade	Math No. of teachers	Math Reliability	Reading No. of teachers	Reading Reliability
K	123	.80	129	.86
1	133	.86	152	.79
2	150	.81	159	.58
3	136	.80	148	.57
4	116	.75	112	.56
5	107	.85	106	.56
6	45	.90	68	.51
7	38	.81	69	.44
8	37	.84	55	.47