Title

Performance Information and Personnel Decisions in the Public Sector: The Case of School Principals

Authors

Cullen, Julie Berry
Hanushek, Eric A
Phelan, Gregory
et al.

Publication Date

2016-12-01

Peer reviewed

Performance Information and Personnel Decisions in the Public Sector: The Case of School Principals[1]

Julie Berry Cullen[2], Eric A. Hanushek[3], Gregory Phelan[4] and Steven G. Rivkin[5]

November 20, 2016

ABSTRACT

Firms and other organizations establish the criteria under which employees will be judged and the performance measures made available to supervisors, the board of directors and other stakeholders, and these structures almost certainly influence behavior and organization outcomes. Any divergence of the chosen performance metric from an ideal measurement of productivity may lead to suboptimal outcomes, particularly in the public sector where outside interest groups may rely more heavily on easily accessible ratings than better-informed insiders. In the case of public education, federal and state accountability systems provide considerable information about student outcomes and rate schools on that basis. However, the No Child Left Behind accountability legislation's focus on pass rates rather than learning and achievement growth introduces the possibility that inadequate information and a flawed structure each compromise public school quality. This study of school principal labor market outcomes investigates the relationship between principal labor market success and a set of performance measures that differ on the basis of accessibility to stakeholders and link with true principal productivity. The results from the empirical analysis provide evidence that information and design deficiencies introduce a lack of alignment between incentives and principal productivity and adversely affect the quality of education in Texas public schools.

[1] This work was done in conjunction with the Texas Schools Project at the University of Texas at Dallas. The conclusions of this research do not necessarily reflect the opinions or official position of the Texas Education Agency, the Texas Higher Education Coordinating Board, or the State of Texas.
[2] University of California, San Diego and NBER
[3] Stanford University, University of Texas at Dallas, and NBER
[4] University of Texas at Dallas
[5] University of Illinois at Chicago, University of Texas at Dallas, and NBER

1. Introduction

Firms and other organizations establish the criteria under which employees will be judged and the performance measures made available to supervisors, the board of directors and other stakeholders, and these structures certainly influence behavior and organization outcomes. The standard principal-agent problem constitutes one channel for suboptimal outcomes, as information failures permit the incentives of the agent to deviate from those of the principal. Divergence of the chosen performance metric from an ideal measure of productivity provides additional channels for suboptimal outcomes, particularly in the public sector where outside interest groups may rely more heavily on the metric than better-informed insiders.

Public schools offer an excellent setting for the examination of the effects of information structures on the labor market outcomes of managers, in this case school principals. Since 2002, the No Child Left Behind (NCLB) accountability legislation requires states to measure and disseminate information on school performance and to establish ratings and sanctions based on these measures. Importantly, the law focuses on pass rates rather than more informative measures of learning. Not only do families and other non-school factors influence the likelihood of passing a test, passing indicators ignore all achievement growth that does not cause a student to cross the passing threshold. Of particular concern is the possibility that the accountability system unfairly disadvantages schools serving high-poverty families, complicating efforts to attract and retain effective school leaders.[6]

The state of Texas, the focus of this analysis, makes overall school pass rates and pass rates for specified demographic groups publicly available online along with a rating that places school performance into four broad categories according to a complicated set of rules. Nonetheless, it is not clear how superintendents make use of such information in personnel decisions. Even if district administrators have extensive knowledge of the pass rates along with other information that goes beyond the publicly prescribed metrics, school boards and parents may fail to access it and instead push for action on the basis of the widely publicized school ratings. Receiving the lowest rating of unacceptable would be expected to elicit the strongest response, potentially leading to very different treatment of nearly identical principals on opposite

[6] Li (2015) finds evidence that NCLB decreases average principal quality at schools serving disadvantaged students in North Carolina, as effective principals seek positions at schools less likely to face sanctions.

sides of the ratings cutoff. On the one hand, this may place pressure on a reluctant superintendent to remove one under-performing principal and leave the other in her current position despite the fact that both should be removed. On the other hand, an unacceptable rating may prompt a superintendent to remove an effective principal, since the reliance of stakeholders on the rating may weaken the superintendent's arguments in favor of retention.

Of course some district administrators and school boards may not only make use of more detailed pass rate information but also recognize the weaknesses of the existing accountability system — leading them to establish policies or practices that incorporate additional information more closely associated with school effectiveness into evaluations. One potential measure is the change in the pass rate from the prior year. Because districts must collect the student-level data reported to the state accountability system, they might also calculate performance measures based on achievement growth or value-added or may develop other evaluation systems associated with value-added for use in personnel decisions.

We conduct a two-pronged empirical analysis of the impact of school performance data on principal labor market outcomes. The first component examines associations between school performance and a series of labor market outcome measures for principals, comparing patterns for the accountability-specified measures of performance with measures that are arguably more closely related to principal effectiveness, such as the change in student pass rates over time and an estimate of school value-added to achievement. We account for observable characteristics of students and principals, but the possibility of unobserved school and principal factors confounding the estimated effects of the performance measures inhibits drawing causal inferences based on these regressions.

The second component of the empirical analysis uses regression discontinuity design methods to identify the impacts of crossing a rating threshold. As long as no considerable sanctions or rewards are triggered, the rating itself should have little effect on labor market outcomes in the absence of an information failure. Since principals leading schools that fall barely on either side should not be differentiable in terms of effectiveness, any significant ratings effect at the boundary constitutes evidence that some stakeholders do not make use of the detailed information underlying the ratings. In the context of home prices and residential location, Figlio and Lucas (2004) find evidence of this type of ratings effect, suggesting that homebuyers rely at least in part on the readily available and salient ratings despite the

availability of comprehensive information on the components. If school administrators, who are presumably far more knowledgeable about the evaluation system, also treat equally productive principals differently, it suggests that the structure of accountability regulation can to some degree interfere with school governance.

Both components suggest that information failures do in fact lead to coarse and uneven treatment of principals. From the first analysis, principal labor market outcomes are strongly related to overall student pass rates and to school ratings but show little or no relationship with the more precise measures of school effectiveness found in changes in pass rates or in school value-added. In the regression discontinuity analysis, receipt of an unacceptable rating, even if it is not accompanied by formal sanctions on districts, leads to substantially worse labor market outcomes relative to principals whose schools are just across the acceptable threshold. The contrast between the very small and insignificant effects of crossing the recognized and exemplary thresholds and the large and significant effect of not escaping the stigmatizing unacceptable rating suggests that it is not administrators but rather school board members, families, or other stakeholders not directly involved in school management who fail to access or use detailed performance information. Taken as a whole, the results show that both the accountability system structure and stakeholder use of information influence the careers of school leaders, providing another mechanism for school accountability to influence the operation and effectiveness of public schools. Of particular importance, it appears to disadvantage principals of higher-poverty schools by permitting outside factors to reduce the probability of labor market success.

These results extend the few existing prior studies of the principal labor market reward structure. In prior work on Texas, Cullen and Mazzeo (2008) find that first-time principals who lead schools where achievement is higher than expected given the student demographics are more likely to move to more advantaged schools and to be promoted, realizing larger salary increases through these channels. In comparison to that analysis, we use student level data, extensive information on accountability ratings, and regression discontinuity methods to compare the market response associated with more and less salient measures of principal effectiveness.

The next section describes the Texas administrative data, and Section 3 provides relevant background on the Texas school principal labor market and school accountability system.

Sections 4 and 5 detail the measurement of principal effectiveness and labor market success, respectively. Section 6 describes the association between various measures of principal effectiveness and labor market success. Section 7 presents causal estimates of the effects of crossing ratings thresholds based on regression discontinuity analyses. Section 8 explores the possibility that the effects may vary by district size due to differences in personnel policies and practices. Finally, Section 9 summarizes the findings and considers implications for policy.

2. School and Labor Market Data

In order to characterize labor market outcomes for principals and the performance of the schools they lead, we use a combination of restricted-use and publicly available data. We focus on principals leading elementary schools over the 2001 to 2008 school years, where school years are identified by the spring years. The focus on elementary grades allows us to summarize a school's academic success using achievement metrics, whereas the broad range of attainment outcomes increases the dimensionality in later grades. For the early grades, achievement tests have been consistently administered for consecutive grades, making it is possible to observe achievement growth in addition to achievement levels. The choice of the sample period is driven by data availability.

The restricted-use data we rely on is the administrative data constructed as part of the UTD Texas Schools Project.[7] Working with the Texas Education Agency (TEA), this project has combined different data sources to create matched panel datasets of staff and students. The personnel database reports annual information on administrator background characteristics, experience both as a teacher and administrator, and current position and salary. From this information we are able to accurately track the careers of principals as long as they remain in Texas public schools. The student panels include demographic characteristics, instructional program participation, and achievement test scores. Using campus identifiers in the panel datasets, we are able to merge data from the publicly available Texas Academic Excellence Indicator System (AEIS). These comprehensive annual reports include a broad range of school-specific contextual and performance measures.

One of the advantages of studying Texas is the large number of principals and schools that are observed. Over our period, there are 3,259 rated elementary schools serving an average

[7] https://www.utdallas.edu/research/tsp-erc/

of 565 students each year. Further, the typical elementary school experiences a principal turnover every 4.3 years.

3. Institutional Background

The principal labor market in Texas is likely to be fluid relative to other states. Texas is one of the few states that prohibit public employees from entering into collective bargaining. School principals and teachers generally serve under term contracts, and those contracts cannot be longer than five years and are typically shorter. Principals come from the teacher ranks, as they are required to have two years of classroom teaching experience in addition to completing a Master's degree in a principal preparation program. Although there is a state minimum salary schedule for teachers by years of experience, there are no constraints on principal salaries. Salaries for principals are set by the superintendent and subject to approval of the school district board. For our sample of elementary school principals, the average salary (in 2008 dollars) is $74,979, and ranges from a low of $35,191 to a high of $128,479.

As the school's leader, the principal is responsible for how the school functions. In Texas, principals are required to be evaluated annually by central administrators. State code recommends standards for evaluating principals on specific indicators in the areas of instructional leadership, human capital development, executive leadership, school culture and strategic operations. Importantly, academic progress of students at the school becomes a factor starting in the second year after a principal has been at a campus.

The evaluation of principals takes place within the broader system of statewide standardized testing and school accountability. The accountability system determines not only the publicly available information on school academic outcomes but also the data available to construct additional measures of principal productivity. Texas has required statewide testing since 1980 and was also an early mover on school accountability, having implemented a four-tiered system in 1994. School ratings of unacceptable, acceptable, recognized, and exemplary have been assigned by the state every year since then, with the exception of 2003 due to the transition to a new standardized-testing regime.[8] Elementary school ratings depend primarily on

[8] The Texas Assessment of Academic Skills (TAAS) was administered each spring to students enrolled in grades three through eight starting in 1993. In 2003, this test was replaced by the Texas Assessment of Knowledge and Skill (TAKS). Both are criterion-referenced tests that assess student mastery of grade-specific subject matter.

performance in mathematics and reading, as well as writing and science in grades 4 and 5, respectively.

The mapping from test scores to the campus rating is complex. First, separate pass rates for each subject based on year-specific cutoff scores for proficiency are calculated for all students and for demographic subgroups (black, Hispanic, white and low-income) that meet minimum size requirements ranging from 30 to 50 students. Then, these pass rates are compared to thresholds that vary by rating category and year. The lowest pass rate across subjects and subgroups is the primary determinant of the accountability rating, but there are a number of exceptions. For example, for an acceptable rating, a subgroup not reaching the current statutory threshold in a subject but closing a specified percentage of the gap from the prior year can meet the alternative standard of required improvement.[9] The required improvement alternative is also available for the recognized rating, with the additional requirement that the pass rate fall no more than five percentage points below the statutory rate. The 2004 through 2008 accountability systems also include additional exceptions provisions for campuses to be elevated to acceptable, recognized, and exemplary ratings: a specified number of subject-by-subgroups can be ignored as long as the pass rate falls no more than five percentage points below the statutory rate and the subject-by-subgroup did not receive an exception in the prior year.

Over the years 2001 to 2008, 17 percent of elementary school campuses were rated exemplary, 44 percent were rated recognized, 38 percent were rated acceptable, and only 1 percent received an unacceptable rating. The campus ratings and underlying student performance indicators are linked to both rewards and punishments. The state appropriates limited funding to provide financial awards to schools rated acceptable or above that show sustained or improved performance, as well as to schools led by principals identified as high-performing based on the same types of indicators. The highest performing campuses are also exempted from specific regulations. On the other hand, schools rated as unacceptable must develop improvement plans along with external review teams in the first year. Receipt of an unacceptable rating in two consecutive years initiates the imposition of sanctions that become progressively more severe for each additional year the school fails to reach an acceptable rating.[10] After five years,

[9] In this case, the prior year pass rate is adjusted to account for any change in the cutoff score for passing.
[10] Starting in 2004, when the federal No Child Left Behind policy became effective, schools are also classified by whether they meet adequate yearly progress (AYP). The state aligns that determination as closely as possible to the school rating process, though federal rules require adjustments to some of the indicators considered, including the

requirements to replace school staff or make other dramatic changes can directly affect principal job retention.

All of the detailed and summary information about academic performance is made publicly available on the web. In evaluating principals, district administrators surely have further information to go by, such as measures of performance on other dimensions, teacher reports, feedback from students and families, and direct observations. Yet, the extent to which these sources of information guide personnel decisions might be moderated by pressure from less informed stakeholders. In the next section, we discuss alternative proxies for principal effectiveness and how they differ in salience and observability, as well accuracy.

4. Measures of Principal Effectiveness

A natural way to judge principal effectiveness is by the academic performance of students at the school she leads. However, just as in the case of corporate CEOs, performance depends on many factors that are not directly within the principal's control, including the composition of the student body. In this section, we first describe how we construct a value-added measure of effectiveness and then contrast it to more readily available metrics.

4.1 Principal value-added to student achievement

We start with what we view to be the most convincing measure of principal effectiveness, which is the value-added of the school to achievement. This measure is designed to separate the influences of the school from those of outside factors, including the parents. The estimation of school value-added parallels the more fully developed estimation of teacher value-added but has a number of differences that affect the analytical complexity and interpretation.

The residential location and school choice decisions of families in combination with school assignment policies and practices introduce substantial variation in student composition across schools that must be addressed in studies of both principals and teachers. At the same time, the widely discussed problems for the estimation of teacher value-added associated with purposeful allocation of students to classrooms and test measurement error are far less important in the case of principals given the focus on school-wide performance and the much larger

consideration of additional subgroups. During our sample period, among those elementary campuses designated as failing to meet AYP, only 8 percent were also rated as unacceptable.

number of test-takers in schools than classrooms.[11] However, the persistence of principal influences on the quality of instruction in years after departure presents serious hurdles to the identification of principal effectiveness.

Many actions including teacher hiring, contract renewal decisions, teacher mentoring and the establishment of a school climate will affect the quality of instruction beyond a principal's tenure. This contrasts with the case for teachers, where the longer-term effects on achievement can be captured by lagged achievement measures for observations in later years, and the teacher in the previous grade generally has little or no involvement with instruction in the current year. Even if lagged test scores do not fully account for prior teacher effects due to the dynamics of learning, it is possible to account directly for prior teacher effects in the model.[12] In the case of principals, however, it is clear that prior achievement by itself does not account for effects of decisions that directly affect learning in future periods.

The value-added model used in this paper relates mathematics achievement (A) for student i in grade g in school s in year t to a cubic polynomial in prior mathematics achievement ($f(A_{it-1})$), observed student characteristics (X), time- and grade-varying school and peer characteristics (C), year-by-grade indicators (δ) and a vector of school-by-year fixed effects (γ).[13] Adding a random error (ε), the empirical model is:

$$A_{igst} = \alpha_1 f\left(A_{it-1}\right) + \alpha_2 X_{it} + \alpha_3 C_{gst} + \delta_{gt} + \gamma_{st} + \varepsilon_{igst}$$

The vector X includes indicators for student race and ethnicity, eligibility for subsidized lunch, Title I status, special education participation, limited English proficiency and gender. It also includes indicators for students who change campuses between the fall census date and spring standardized testing. The vector C includes the averages of these demographic characteristics for students in grade g in school s in year t. Our estimate of principal value-added is based on the school-by-year fixed effect (γ).

Recent evidence in Miller (2013) reveals a systematic decrease in school value-added in the year prior to the arrival of a new principal. This may reflect a reduction in principal health,

[11] Hanushek and Rivkin (2010), Chetty, Friedman, and Rockoff (2014), Kane et al. (2013), and Rothstein (2010) investigate the presence and magnitude of biases introduced by nonrandom assignment to classrooms.
[12] Rothstein (2010) shows a relationship between previous teacher quality and achievement even in a value-added specification.
[13] While the general concept has been used in education for over three decades (see Hanushek (1979)), the recent addition of extensive administrative databases has led to expansion of both the empirical analysis (Hanushek and Rivkin (2010)) and the understanding of underlying estimation and interpretation issues (Meghir and Rivkin (2011)).

effort, or authority over the school or the impacts of other factors associated with the decision to leave. Because of the possibility that value-added in a principal's first year might be inflated by a recovery from the achievement dip in the final year of the prior principal's tenure as well as the fact that the persistent influences of the prior principal are likely to be strongest during the first year of a spell, we exclude the first year of job spells from the sample.

The validity of value-added estimates as measures of productivity is a central issue, and two recent papers raise concerns about its use. Grissom, Kalogrides, and Loeb (2015) and Chiang, Lipscomb, and Gill (forthcoming) use sophisticated and data-intensive models to compare value-added estimates to alternative measures of principal quality and raise doubts about the attribution of value-added estimates to the principal. The former paper finds a weak relationship between value-added and the district evaluations of principals, and the latter finds a weak relationship between value-added and estimates of persistent school quality. Importantly, these papers rely heavily on schools with multiple principals who serve short terms as leaders, and both include the initial and final years of principal terms in the analysis. This is particularly worrisome in the estimation of school improvement under each principal in Grissom, Kalogrides, and Loeb (2015) and in models including both school and principal fixed effects in Chiang, Lipscomb, and Gill (forthcoming).

In contrast, Laing et al. (2016) find a strong relationship between value-added and teacher survey responses on principal effectiveness in a similar study that excludes the initial and final year of principal terms. Average value-added increases monotonically with teacher ratings for three questions about the principal as an instructional leader. Moreover, in specifications that include school fixed effects, a regression of school value-added on a teacher rating index based on factor analysis produces a positive and highly significant relationship between the two that is very similar to that produced by a specification that does not include school fixed effects. These findings support the view that value-added estimates capture meaningful variation in principal productivity.

4.2 Accountability-based proxies for principal effectiveness

Since the state prescribes a set metrics of school performance based on student pass rates for the state achievement tests and makes them available online, these are candidates for evaluative information employed by the district and public. We consider three of these pass rate metrics: i) the state rating assigned to the school; ii) the average pass rate across reading and

math subjects; and, iii) the change in the average pass rate from the prior year. The latter is closest conceptually to value-added and captures whether or not performance at the school is improving. Of course, it focuses on improvements just around the cut score for passing and not elsewhere in the distribution, and it absorbs any changes in ability or characteristics of students across cohorts. The annual pass rate is presented in school reports and does not have to be calculated, but it also makes no adjustment for demographics or prior achievement. The school rating is arguably the most salient and widely publicized but also the least informative, since it heavily weights the minimum across a varying number of pass rate and pass rate growth measures.

These alternative measures are imperfectly correlated and do not give a consistent picture of school performance. Table 1 reports the 10^{th}, 25^{th}, 50^{th}, 75^{th} and 90^{th} percentiles of the distributions of value-added and the campus pass rate by state rating category. Although value-added increases monotonically from unacceptable to exemplary at each percentile, there is substantial overlap across the ratings categories. For example, the 50^{th} and 75^{th} percentiles of value-added for unacceptable, acceptable, and recognized exceed the 25^{th} and 50^{th} percentiles, respectively, for the next higher ratings. Not surprisingly, the bottom panel shows that the association between the campus rating and pass rate is far stronger.

Table 2 reports the correlations among value-added, the pass rate, and the change in the pass rate from the prior year for both the reported pass rates and pass rates adjusted for student demographic characteristics. The correlation between the pass rate and value-added increases from 0.31 to 0.44 once the pass rate is adjusted for student composition. The pass rate conditional on demographic characteristics almost certainly provides a better measure of school effectiveness than the raw pass rate. Reinforcing this, the correlation between the pass rate and the change in the pass rate shifts from negative to positive following the adjustments.

Because nonschool factors account for a larger portion of the variation in the more easily observable proxies, this raises the possibility that selection into a school serving higher-SES students may be more beneficial to a principal's labor market prospects than raising the quality of instruction. For the same reasons, it might be difficult to attract principals to a school that is likely to receive a low rating due to limited family resources, for fear of being penalized for any failure. Rating schools based on better measures of student learning could align principal labor market opportunities more closely with productivity as a school leader. Our empirical analysis is

designed to shed light on the signals the market seems to respond to as the system is currently structured.

5. Measures of Principal Labor Market Success

We observe the annual labor market transitions of principals, but we lack direct information on either the choice set or the preferences of each principal. A principal may choose to move to another school or district due to unobserved pull or push factors, so that voluntary and involuntary mobility and lack of mobility are hard to distinguish. Salary on its own is a noisy measure of success since districts may have set salary schedules or multi-year contracts, and working conditions exert substantial influence on principal preferences for schools and districts. There is evidence, for example, that high levels of achievement and advantaged student bodies are among the most important draws for principals and other educators (Loeb, Kalogrides, and Horng (2010), Hanushek, Kain, and Rivkin (2004)).

We employ several measures of success but emphasize a composite indicator of labor market success that incorporates the multiple considerations above. This composite equals one for a principal who either retains her job or makes a "positive" move. To identify positive moves, we consider trajectories for both salaries and working conditions. To be classified as gaining in terms of salary, the mover must experience salary growth that exceeds the median for all principals who remain in the Texas public schools in the subsequent year, regardless of position or location. To be classified as gaining in terms of working conditions, the principal must move to a new principal position where the predicted change in achievement based on the characteristics of students at the school and the overall performance of the district exceeds the median for all principals who remain principals.[14] We also separate positive outcomes into within-district success and out-of-district success in order to learn more about possible differences in information used by current as opposed to potential future employers. Finally, we use salary growth by itself as an alternative measure of success, though principals who exit the Texas public schools must be excluded from the sample.

[14] More specifically, separately by year, we regress the average of the campus reading and math pass rates on average student characteristics (the shares economically disadvantaged, classified as ESL, classified as gifted and talented, classified as special education, and moving campuses during the year), the log of enrollment, and district fixed effects. Observations are weighted by enrollment. The predicted values from this regression are then standardized to have a mean of zero and standard deviation of one.

The residual categories of principals who are identified as not being successful include principals who move to lower paying and less appealing positions within the district, as well as principals who exit the public school system. This latter category is quite heterogeneous. Individuals who exit may be switching to private schools, changing occupations, dropping out of the labor force, or retiring. Though we are unable to differentiate among alternative destinations, we attempt to reduce the share who retire by excluding principals with more than 25 years of total experience as a teacher, principal, or other school professional employee.

An important issue to consider when linking success to our measures of school performance is timing. While preliminary results are available to district officials as early as May, the final accountability ratings and underlying student achievement data are released annually in early August. Given that most principal hiring occurs in the spring, there is limited scope for immediate impacts on principal positions in the subsequent fall. We therefore use a two-year definition of success, relating labor market outcomes in academic year $t+2$ to performance as measured in the spring of academic year t.

Our sample thus starts with all principals with 25 or fewer years of experience leading elementary schools during the school years 2001 to 2008. We then exclude those cases where the principal is in the first year of a spell at a school – both because of the difficulties noted in estimating value-added in the first year and because new principals may have little influence over the stock of teachers, school climate, and school practices during their first year. We observe 4,241 unique elementary school principals and 11,428 principal-by-year labor market transitions.

Table 3 shows the probabilities of making the different two-year labor market transitions overall and by campus rating. The rate of composite success is 84 percent, with only a small share attributable to across district moves. The rate of composite success rises with the rating, increasing from 51 percent for principals in schools rated unacceptable to 81 percent in schools rated acceptable and to more than 85 percent for those rated recognized or exemplary. Column 2 illustrates that success within the district, primarily retention in the same principal position, follows the same monotonic pattern. Rates of out-of-district success among those who do not have within-district success do increase with rating, even though the raw rates of out-of-district success decrease with the rating in Column 3. Finally, the rate of salary change is positively association with rating, with those in schools rated unacceptable who remain in the Texas public

schools suffering an average salary decline that exceeds three percent.

Positive associations between the probability of success and the other school performance measures also appear (not shown). For example, the pass rate is approximately three percentage points higher and value-added is roughly 0.04 standard deviations higher for principals who succeed. In the next sections, we more formally evaluate the relative roles of the alternative performance measures in moderating labor market outcomes.

6. Principal Effectiveness and Labor Market Success

We conduct two types of empirical analyses. The first, described in this section, studies the associations between principal labor market success and the several proxies for effectiveness. The second, described in the next section, uses a regression discontinuity design to identify the causal impact on labor market success of crossing a ratings threshold.

Table 4 reports the results from three ordinary least squares specifications each for the composite success and salary change outcomes. The first includes the accountability rating and pass rate variables, the second substitutes value-added for the pass rate measures, and the third includes all measures together. The table reports robust standard errors clustered by district. Also, in order to focus on the role of information, all specifications include student and principal characteristics as well as year fixed effects.[15] The controls include indicators for principal ethnicity, degree level, gender and tenure, and proportions of students who are black, Hispanic, white, Asian, economically disadvantaged, classified as ESL, classified as gifted and talented, and classified as special education.

The first three columns of Table 4 illustrate the strong association between labor market success of the principal and both the average school pass rate and the accountability ratings. A significant relationship between the probability of composite success and value-added emerges only for specifications that exclude the pass rate, and the change in pass rate never enters significantly. These results suggest that districts primarily use performance information that is readily available online as opposed to information more closely related to value-added. In addition, the highly significant ratings coefficients conditional on the pass rate suggest that

[15] The qualitative findings are robust to including district fixed effects as well (in results that are not shown).

ratings have independent effects.[16] Receipt of an unacceptable rating is associated with lower rates of success by about 25 percentage points, while receipt of a recognized or exemplary rating is associated with a much smaller 2-4 percentage point gain.

Columns 4-6 report results for the same set of specifications estimated with the rate of change in salary as the dependent variable. Similar to the composite success specifications, the relationship is strongest with the pass rate and the unacceptable rating. In contrast, the estimated coefficients on value-added are small and insignificant regardless of whether the pass rate is included, and the recognized rating loses statistical significance when the pass rate is included.

The decisions of both the current district and potential employers determine the probability of labor market success, leading us to examine the relationship between each of the components of our success indicator and the school performance measures. Current districts are almost certain to have better performance information than other potential employers, suggesting that the probability of continued employment or increases in compensation in the current district may be more strongly related to value-added than would the probability of a successful transition to another district. By comparison, potential employers might be expected to rely more heavily on more readily available information including pass rates and ratings. But, even if the current superintendent has extensive knowledge about school and principal performance, she may well face pressure from stakeholders to take action in the case of low pass rates or, more specifically, an unacceptable rating. Consequently, it would not be surprising to find that an unacceptable rating substantially reduces the probability of success within the district.

Table 5 reports multinomial logit estimates that separate within- and out-of-district success for the same progression of controls found in Table 4. The average pass rate positively affects both within- and out-of-district labor market success of the principal, while conditional on the pass rate, there is no evidence of a significant relationship between value-added and either type of success. Moreover, the school accountability rating only has a significant relationship with success within-district, consistent with the hypothesis that the school ratings help to shape the opinions of the parents, school board, and other interested local people.

In general, though, the differential impacts should not be over-interpreted. The out-of-district success estimates are too imprecise to draw strong conclusions regarding differences

[16] Though controlling for the pass rate linearly is potentially restrictive, the inclusion of higher-order pass-rate terms does not attenuate the ratings coefficients.

between internal and external labor markets. This is due to the relatively small number of principals who realize a successful move to another district.

7. Causal Impacts of Ratings on Success

The results from the preceding section raise the possibility that districts, including the current employer, make use of ratings for personnel decisions despite the availability of detailed information on pass rates. But these estimates provide far from convincing evidence that ratings actually affect labor market outcomes, conditional on the measures that determine those ratings. First, though we control for the overall campus pass rate, we have not controlled for the performance of subgroups or potential nonlinearities in the effects of pass rates on personnel decisions. Second, employer and principal perceptions, preferences and characteristics may differ systematically by factors that are correlated with the campus rating.

In order to isolate the causal effects of ratings, we use regression discontinuity design (RDD) methods based on the school accountability system rules. RDD provides an ideal method for addressing the information question by providing estimates of the effect of a higher rating for otherwise identical principals in terms of underlying performance.

7.1 Regression discontinuity design

Our RDD exploits discontinuities in the probability of receiving a higher accountability rating based on the pass rate for the subgroup (i.e., student group *x* subject) that is most likely to be binding for that campus and year. In order to construct the running variable, we must identify this marginal subgroup. For each rating boundary, we first determine the relevant pass rate threshold for each subgroup that meets applicable minimum size requirements. The threshold may be the statutory threshold, the required improvement threshold, or the exceptions threshold and is determined by the subgroup pass rate in the prior year and whether exceptions are available. We then center pass rates around their relevant thresholds for each of the three rating boundaries. The subgroup with the most negative (or least positive) centered pass rate is selected as the marginal subgroup for each rating category. Running variable values greater than (less than) zero indicate that student performance was sufficient (not sufficient) to earn the higher rating.

We estimate our models using local linear regression with a triangular kernel.[17] We use the structure of the accountability system and existing research to guide our choice of bandwidth. The distances between the statutory pass rates for the various ratings leads us to trim the samples to schools with running variable values within ten percentage points of the threshold in question. Virtually all schools within this range earn one of the two ratings around the threshold, while the fraction falling into a different rating category rises rapidly outside this range. We apply two alternative bandwidths to the trimmed sample—the Imbens-Kalyanaraman optimal bandwidth and the full sample of schools within ten percentage points of the threshold. We cluster standard errors by values of the running variable in all specifications.

Figure 1 illustrates the relationship between the probability of attaining a higher rating and the running variable for each of the thresholds. The discontinuity is quite pronounced at all three thresholds. Though we fully incorporate the complex rules that change over time in the construction of the running variable, the presence of a small fraction (less than 2 percent) of schools whose ratings we do not correctly predict means that we have a fuzzy design.[18] The corresponding first-stage estimates reported in Appendix Table A1 range from between 0.82 and 0.88 at the unacceptable-acceptable boundary, whereas they all exceed 0.96 at the recognized boundary and 0.94 at the exemplary boundary. Consequently, though we report intention-to-treat estimates for the labor market outcomes, local average treatment effect estimates are not too different in magnitude.

Any discontinuities in outcomes at the thresholds can be attributed to the receipt of the rating only if principals are unable to manipulate the running variable near the boundary and no other determinants of outcomes differ discontinuously at the boundary. Though others have shown that it is possible to manipulate pass rates by altering the test-taking pool (Cullen and Reback (2006)), it is not feasible to do so precisely. Once students sit for exams, exam documents are scored and recorded centrally, so that variation in the subgroup pass rates in the neighborhood of the thresholds should be as good as random. Figure 2 shows the densities of acceptable, recognized, and exemplary running variables. The densities for acceptable and exemplary trend relatively smoothly through the threshold, as would be expected. There is a noticeable jump up at the threshold for recognized that we are presuming is the result of chance.

[17] Rectangular kernels produce very similar estimates.
[18] The accountability manual indicates that accommodations may be made in particular circumstances that are not elucidated.

To reinforce this interpretation, we test whether the samples of schools are balanced in terms of observable characteristics on either side of the ratings thresholds. We estimate a system of seemingly unrelated regressions using principal and student population characteristics as the dependent variables.[19] Appendix Table A2 shows that few of the covariates exhibit statistically significant discontinuities at the ratings boundaries. We fail to reject the null hypothesis that all coefficients are jointly equal to zero for both the acceptable and recognized boundaries, but we do reject the joint equality hypothesis at the exemplary boundary. This rejection reflects some relatively small differences: the white student share is estimated to decrease by 4 percentage points while the disadvantaged student share increases by the same amount just above the threshold; principals above the exemplary threshold are estimated to have 0.3 years greater tenure. None seem to suggest much positive selection. Regardless, the finding that labor-market outcomes including salary change move smoothly through the exemplary threshold illustrated below provides evidence of the absence of manipulation by the principal in an effort to further her career.

7.2 RDD estimates of the impacts of school ratings

We estimate the impact of barely attaining the next highest rating on three labor market outcomes – composite success, within district success, and rate of change in salary. Figures 3-5 plot the relationships between the running variables for each of the ratings boundaries and these outcomes. Tables 6 and 7 report the RDD estimates for the two selected bandwidths.

For the recognized and exemplary thresholds, there is little evidence that the campus rating affects labor market outcomes. Each plot in Figures 4 and 5 moves smoothly through the threshold, and the estimated discontinuities for composite success, within district success and the change in salary are close to zero and not significant at conventional levels. It is important to note, though, that the precision of the estimates is not great enough to rule out small effects on the order of those found in the prior section.

In contrast, large positive impacts of crossing the acceptable threshold on success and salary are clear in Figure 3. Corroborating the plots, the top row of Table 6 reports optimal bandwidth estimates of impacts on the labor market outcomes, and all three coefficients are positive and significant at the five percent level. Coefficient magnitudes suggest that crossing the

[19] Principal characteristics include sex, race/ethnicity indicators, years of tenure in current position, years of experience in Texas public schools, and level of degree. School population characteristics include student shares by race/ethnicity, economic disadvantage, gifted and talented, and special education.

threshold raises the overall probability of success by almost thirty percentage points, the probability of success within the district by almost 40 percentage points, and the rate of salary increase by 4.5 percentage points. A closer look at transitions highlights the interrelationships among these outcomes: principals with an acceptable rating are 14.7 percentage points less likely to return to lower-paying teacher or assistant principal positions than principals in schools rated unacceptable.

The top row of Table 7 reports corresponding estimates based on the wider bandwidth of ten percentage points around the threshold. Given the larger sample and range, it is not surprising that the standard errors are somewhat smaller. The coefficients are also smaller, though, so that there is no longer a statistically significant discontinuity for the change in salary. The results are otherwise qualitatively similar, supporting the conclusion that an unacceptable rating adversely affects labor market outcomes.

An important issue concerns the channels that underlie the ratings effect. The regulatory link between sanctions and an unacceptable rating raises the possibility that statutory requirements rather than administrator discretion lead to the observed unacceptable rating effects. However, it takes two unacceptable ratings in successive years to trigger sanctions, so that schools not classified as unacceptable in the prior year are not at risk for sanctions. Table 8 shows that the RDD estimates for these schools are sizeable and significant, reinforcing the conclusion that school ratings provide information that influences discretionary personnel decisions.

The contrast between the strong effects of crossing the acceptable threshold and minimal effects of crossing the recognized or exemplary thresholds raises questions about the sources of the divergent findings. Because administrators in Texas have extensive knowledge about the complexities of the accountability system, it is likely that most understand that variation in the number of subgroups, student demographics, and other factors outside the control of the principal affect the rating. Moreover, as evidenced by the smoothness at the recognized and exemplary boundaries, administrators most likely also use information on pass rates to make finer distinctions about performance than the four categories of the accountability system. Consequently, we believe that the most plausible explanation concerns the failure of the school board or other stakeholders to access or use the more-detailed and more reliable pass-rate information. The stigma of an unacceptable rating is likely to be particularly strong, providing

pressure on the superintendent to terminate the principal. Principals may also understand these dynamics and the additional pressure, and a higher likelihood of a voluntary separation may also contribute to the unacceptable effect.

8. Heterogeneity by District Size

To this point, we have grouped all districts together under the assumption that practices and policies are similar across the state. However, there is good reason to believe that human-resource practices differ among districts. Differences in the capacity to absorb and make use of performance metrics by district size may be particularly important, though this is not a clear cut issue. Larger districts are more likely to have evaluation and assessment specialists and more administrators between the superintendent and principal, and central administrators in larger districts may have less personal contact with families. Although this suggests that larger districts would make greater use of information, it does not imply that they would rely more heavily on ratings, as more capacity to analyze data would be expected to result in more reliance on the detailed information that underlies the ratings. Similarly, superintendents in small districts might receive more input from families and have more personal interactions with principals, but it is not clear that these subjective reviews of quality would be highly correlated with value-added to achievement. Moreover, large districts may establish hierarchical administrative structures that break-up the district into a number of sub-districts in an attempt to create an environment more similar to smaller districts.

In order to learn more about heterogeneity by district size, we divide the sample into thirds on the basis of the number of elementary schools in the district. Table 9 reports mean outcomes and school demographic characteristics for small districts (1-7 schools), medium districts (8-29 schools), and large districts (30-204 schools). Not surprisingly, the likelihood of a successful transition to another district declines with district size, while the probability of a successful outcome within the district rises with district size. The probability of exiting the Texas public schools entirely also declines with district size. In terms of student demographic characteristics, poverty, share black or Hispanic, and share limited English proficient increase monotonically with district size, while there is little systematic variation in the special education or gifted and talented shares.

Tables 10-12 report the estimates from reproducing the empirical analysis for the district size groups. Note that we report ordinary least squares and multinomial logit estimates from the full model only, and we report RDD results only for the acceptable cutoff. Somewhat surprisingly, no strong patterns emerge by district size in either the associations between labor market outcomes and the performance measures or in the RDD estimates of ratings effects. The differences across categories tend to be small relative to the standard errors and do not provide compelling evidence of divergence in personnel practices. In particular, value-added is never significant regardless of outcome or district size. And although crossing the acceptable threshold significantly affects the probability of success and salary change just for the largest districts, the magnitudes of the effects on composite success are actually quite similar in the samples of the largest and smallest districts.

9. Conclusions

The results of this paper provide evidence that information structures affect the principal labor market. The strong association between labor market outcomes and both the average pass rate and school rating illustrates the influence of the accountability system. There is little evidence of a significant relationship with either the change in pass rate or value-added, measures more closely related to school productivity and principal effectiveness. The use of clear school ratings – whether in the form of the Texas categories or in the form of A through F grades for schools – has been advocated to ensure that parents are aware of the quality of their schools, but these results suggest an unintended consequence is their inappropriate use in judging the effectiveness of principals. The contrast between the minimal effects of crossing the higher boundaries and the large effects of crossing the stigmatizing unacceptable boundary is consistent with less informed stakeholders not directly involved in school management applying pressure.

Overall, our results suggest that high poverty schools are likely to be particularly poorly served by the focus on pass rates and associated ratings, since both of these metrics are strongly influenced by student body composition. Holding teachers and administrators accountable for factors outside their control will likely hinder efforts of high-poverty schools to attract and retain effective educators. Nonetheless, the substantial effects of the performance measures currently emphasized in the accountability system suggest that aligning the performance evaluation system better with student learning could improve the quality and allocation of schools leaders. Recent

work on teacher transitions finds that the distribution of teacher value-added information positively influenced personnel decisions and the distribution of teacher quality (Bates 2015).

References

Bates, Michael. 2015. "Public and private learning in the market for teachers: Evidence from the adoption of value-added measures." University of California, Riverside (January).

Chetty, Raj, John N. Friedman, and Jonah Rockoff. 2014. "Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood." *American Economic Review* 104, no. 9 (September): 2633–2679.

Chiang, Hanley, Stephen Lipscomb, and Brian Gill. forthcoming. "Is School Value Added Indicative of Principal Quality?" *Education Finance and Policy*.

Cullen, Julie B., and Michael J. Mazzeo. 2008. "Implicit Performance Awards: An Empirical Analysis of the Labor Market for Public School Administrators." University of California, San Diego (December).

Cullen, Julie Berry, and Randall Reback. 2006. "Tinkering toward accolades: school gaming under a performance accountability system." *Advances in Applied Microeconomics* 14: 1-34.

Figlio, David N., and Maurice E. Lucas. 2004. "What's in a Grade? School Report Cards and the Housing Market." *The American Economic Review* 94, no. 3 (June): 591-604.

Grissom, Jason A., Demetra Kalogrides, and Susanna Loeb. 2015. "Using Student Test Scores to Measure Principal Performance." *Educational Evaluation and Policy Analysis* 37, no. 1 (March): 3-28.

Hanushek, Eric A. 1979. "Conceptual and empirical issues in the estimation of educational production functions." *Journal of Human Resources* 14, no. 3 (Summer): 351-388.

Hanushek, Eric A., John F. Kain, and Steve G. Rivkin. 2004. "Why public schools lose teachers." *Journal of Human Resources* 39, no. 2 (Spring): 326-354.

Hanushek, Eric A., and Steven G. Rivkin. 2010. "Generalizations about using value-added measures of teacher quality." *American Economic Review* 100, no. 2 (May): 267-271.

Kane, Thomas J., Daniel F. McCaffrey, Trey Miller, and Douglas O. Staiger. 2013. *Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment*. MET Project: Bill and Melinda Gates Foundation (January).

Laing, Derek, Steven G. Rivkin, Jeffrey C. Schiman, and Jason Ward. 2016. "Decentralized Governance and the Quality of School Leadership." NBER Wroking Paper No. 22061. Cambridge, MA: National Bureau of Economic Research (March).

Li, Danielle. 2015. School accountability and principal mobility: how No Child Left Behind affects the allocation of school leaders. Harvard Business School Working Paper No. 16-052.

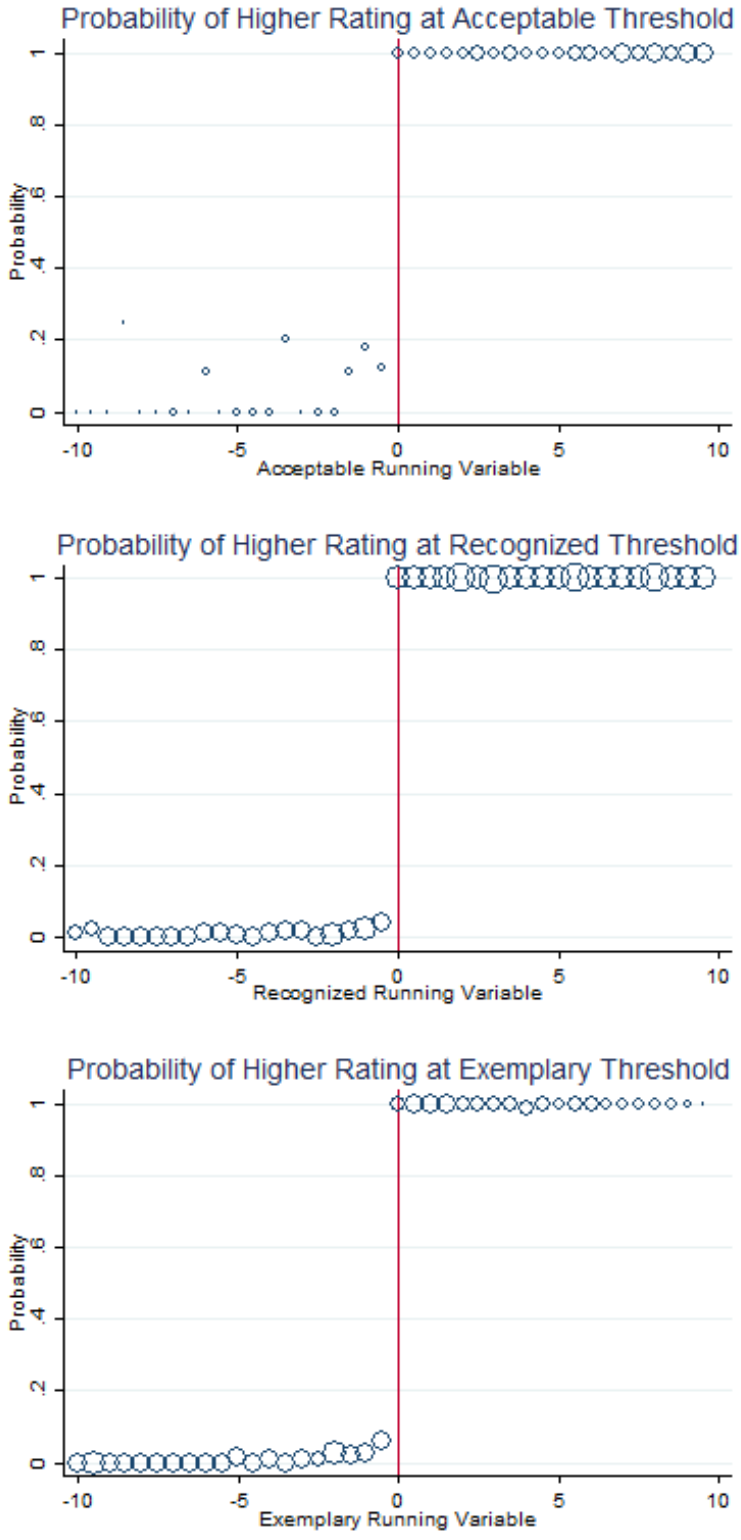
Loeb, Susanna, Demetra Kalogrides, and Eileen Lai Horng. 2010. "Principal Preferences and the Uneven Distribution of Principals Across Schools." *Educational Evaluation and Policy Analysis* Vol. 32, no. 2 (June): 205–229.

Meghir, Costas, and Steven G. Rivkin. 2011. "Econometric methods for research in education." In *Handbook of the Economics of Education, Vol. 3*, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann. Amsterdam: North Holland: 1-87.

Miller, Ashley. 2013. "Principal turnover and student achievement." *Economics of Education Review* 36(October): 60-72.

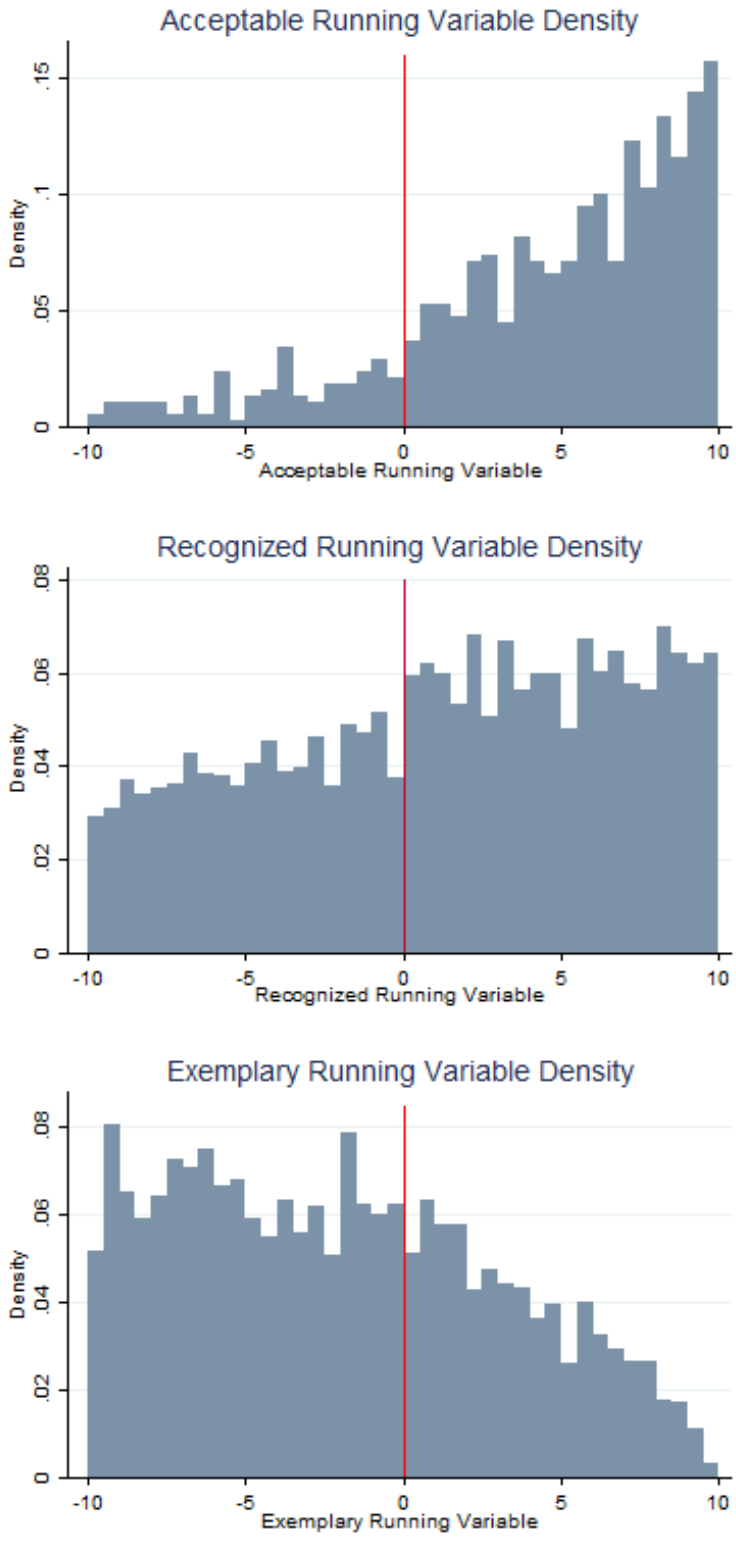
Rothstein, Jesse. 2010. "Teacher quality in educational production: Tracking, decay, and student achievement." *Quarterly Journal of Economics* 125, no. 1 (February): 175-214.

Figure 1: Regression Discontinuity First Stages by Rating



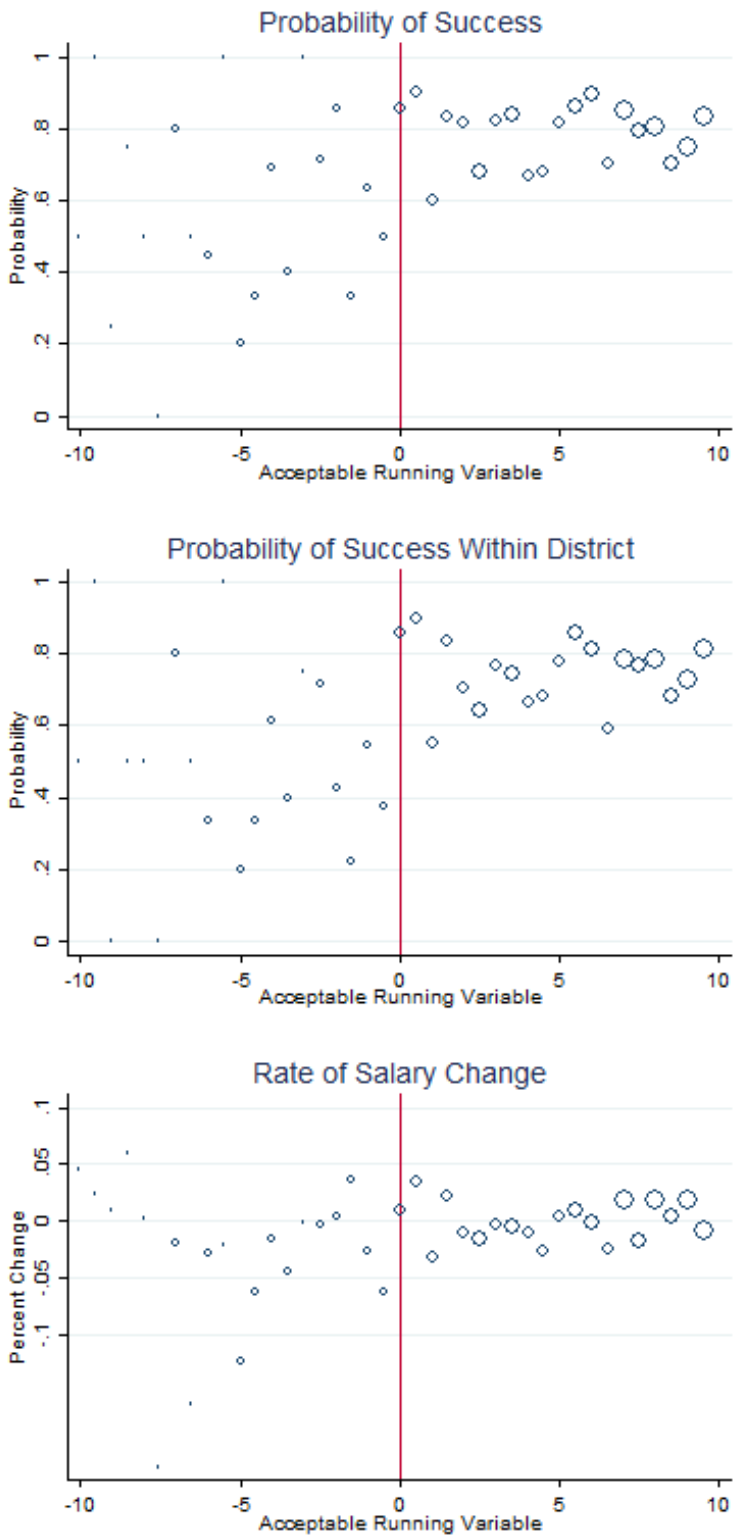
Notes: Bin width = 0.5 percentage points. Points are weighted by bin size.

Figure 2: Running Variable Density by Rating



Notes: Bin width = 0.5 percentage points.

Figure 3: Principal Labor Market Outcomes at Acceptable Threshold

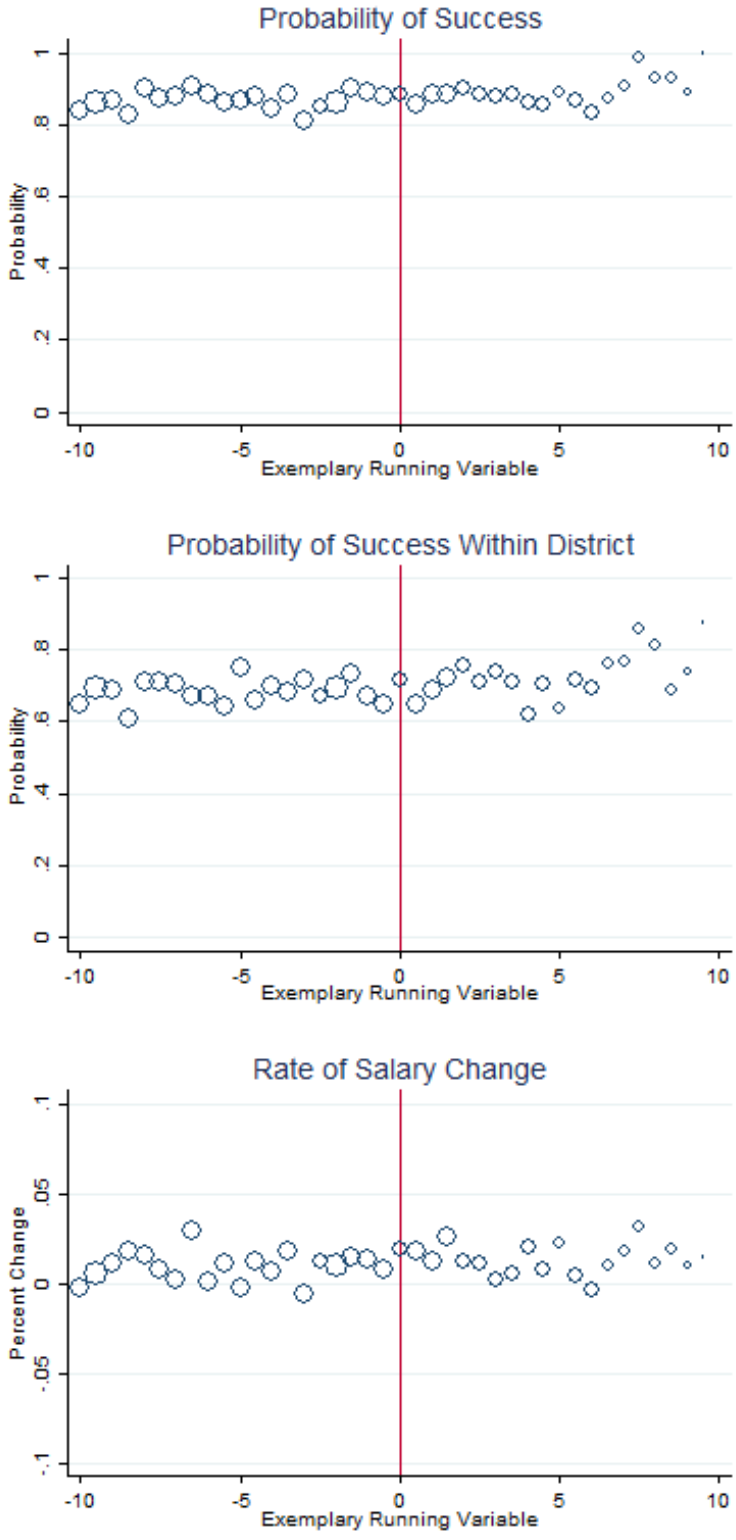


Notes: Bin width = 0.5 percentage points. Points are weighted by bin size.

Figure 4: Principal Labor Market Outcomes at Recognized Threshold

Notes: Bin width = 0.5 percentage points. Points are weighted by bin size.

Figure 5: Principal Labor Market Outcomes at Exemplary Threshold



Note: Bin width = 0.5 percentage points. Points are weighted by bin size.

Table 1: Value-added and Pass Rate at Varying Percentiles, by School Rating

	Percentile				
	10	**25**	**50**	**75**	**90**
Value-added					
Unacceptable	-0.544	-0.357	-0.169	-0.042	0.145
Acceptable	-0.303	-0.176	-0.052	0.072	0.176
Recognized	-0.196	-0.083	0.026	0.131	0.238
Exemplary	-0.096	-0.017	0.067	0.155	0.250
Pass Rate					
Unacceptable	0.555	0.622	0.695	0.769	0.827
Acceptable	0.722	0.775	0.829	0.874	0.909
Recognized	0.843	0.877	0.913	0.940	0.959
Exemplary	0.943	0.958	0.973	0.986	0.994

Notes: Sample includes all principals that had at least 2 years of tenure in the school and 25 or fewer years of total experience in Texas public schools in the school years 2000-01 through 2007-08 (excluding 2002-03 since ratings were not assigned that year).

Table 2: Correlations Among Value-added, Pass Rate, and Change in Pass Rate, by Adjustment for Student Demographic Characteristics

	Value-added	Pass Rate	Change in Pass Rate
Unadjusted			
Value-added	1.00		
Pass Rate	0.31	1.00	
Change in Pass Rate	0.30	-0.06	1.00
Adjusted			
Value-added	1.00		
Pass Rate	0.44	1.00	
Change in Pass Rate	0.37	0.26	1.00

Notes: Adjusted correlations are correlations of the residuals from regressing value-added, pass rate, and one year difference in pass rate on district and year fixed effects and the following controls: indicators for principal ethnicity, degree level, gender and tenure, and proportions of students who are black, Hispanic, white, Asian, economically disadvantaged, classified as ESL, classified as gifted and talented, and classified as special education.

Table 3: Probability of a Successful Labor Market Outcome and Salary Rate of Change, by School Rating

	Composite Success	Within-District Success	Out-of-District Success	Salary Rate of Change
Unacceptable	51.15	42.75	8.40	-0.032
Acceptable	80.80	75.54	5.25	0.002
Recognized	85.90	81.65	4.25	0.009
Exemplary	88.34	84.73	3.61	0.014
All	83.99	79.42	4.57	0.007
Observations	11,428	11,428	11,428	10,529

Notes: Campus ratings are from the spring of academic year *t* for the years 2000-01 to 2007-08 (excluding 2002-03). Composite success is set equal to 1 for a principal who remains in her position or moves to a new position in *t+2* and realizes an above-median salary increase (relative to those who stay in the system) or an above-median increase in predicted student achievement (relative to those who remain principals); it is set equal to 0 otherwise.

Table 4: Estimated OLS Effects of School Performance Metrics on the Probability of Labor Market Success

	Composite Success			Salary Rate of Change		
	(1)	(2)	(3)	(4)	(5)	(6)
Pass Rate	0.491***		0.473***	0.089***		0.086***
	(0.107)		(0.112)	(0.025)		(0.024)
Change in Pass Rate	-0.036		-0.055	-0.019		-0.022
	(0.081)		(0.084)	(0.019)		(0.020)
Value-added		0.053**	0.016		0.009	0.003
		(0.023)	(0.025)		(0.006)	(0.006)
Accountability Rating						
Unacceptable	-0.249***	-0.292***	-0.249***	-0.029***	-0.037***	-0.029***
	(0.035)	(0.035)	(0.035)	(0.011)	(0.012)	(0.011)
Recognized	0.025***	0.051***	0.025**	0.003	0.007***	0.002
	(0.009)	(0.010)	(0.010)	(0.002)	(0.002)	(0.002)
Exemplary	0.037**	0.074***	0.036**	0.007*	0.014***	0.007*
	(0.015)	(0.015)	(0.015)	(0.004)	(0.003)	(0.004)
Observations	11,428	11,428	11,428	10,529	10,529	10,529
R-squared	0.023	0.026	0.027	0.053	0.051	0.053

Notes: OLS estimates with robust standard errors in parentheses. Acceptable is the excluded rating category. All specifications include year fixed effects and the following controls: indicators for principal ethnicity, degree level, gender, tenure, and proportions of students who are black, Hispanic, white, Asian, economically disadvantaged, classified as ESL, classified as gifted and talented, and classified as special education. *** p<0.01, ** p<0.05, * p<0.1

Table 5: Estimated Multinomial Logit Effects of School Performance Metrics on the Probability of Labor Market Success Within District and Out of District

	(1)	(2)	(3)
Within-District Success			
Pass Rate	3.480***		3.363***
	(0.765)		(0.811)
	[0.451]		[0.438]
Change in Pass Rate	-0.472		-0.596
	(0.575)		(0.584)
	[-0.108]		[-0.123]
Value-added		0.368**	0.109
		(0.163)	(0.184)
		[0.042]	[0.013]
Accountability Rating			
Unacceptable	-1.215***	-1.496***	-1.213***
	(0.176)	(0.166)	(0.177)
	[-0.240]	[-0.299]	[-0.240]
Recognized	0.217***	0.394***	0.213***
	(0.069)	(0.072)	(0.069)
	[0.037]	[0.060]	[0.037]
Exemplary	0.360***	0.608***	0.351***
	(0.122)	(0.120)	(0.123)
	[0.054]	[0.082]	[0.054]
Out-of-District Success			
Pass Rate	2.785***		2.642**
	(1.050)		(1.083)
	[-0.001]		[-0.003]
Change in Pass Rate	0.985		0.834
	(0.999)		(1.113)
	[0.056]		[0.056]
Value-added		0.455*	0.130
		(0.273)	(0.321)
		[0.007]	[0.002]
Accountability Rating			
Unacceptable	-0.156	-0.369	-0.154
	(0.320)	(0.307)	(0.319)
	[0.041]	[0.037]	[0.041]
Recognized	-0.090	0.039	-0.094
	(0.132)	(0.127)	(0.132)
	[-0.011]	[-0.012]	[-0.011]
Exemplary	-0.055	0.117	-0.064
	(0.231)	(0.227)	(0.232)
	[-0.014]	[-0.014]	[-0.014]
Observations	11,428	11,428	11,428

Notes: Multinomial logit coefficient estimates with robust standard errors in parentheses. Average marginal effects are reported in brackets for pass rate, change in pass rate and value-added. Differences in probabilities of outcomes are reported in brackets for accountability ratings. Acceptable is the excluded rating category. All specifications include year fixed effects and the following controls: indicators for principal ethnicity, degree level, gender, tenure, and proportions of students who are black, Hispanic, white, Asian, economically disadvantaged, classified as ESL, classified as gifted and talented, and classified as special education. *** p<0.01, ** p<0.05, * p<0.1

Table 6: RDD Intent to Treat Estimates of Rating Effects on the Probability of Labor Market Success, Imbens-Kalyanaraman Optimal Bandwidth

	Composite Success	Within-District Success	Salary Rate of Change
Acceptable	0.294**	0.379***	0.045**
	(0.130)	(0.127)	(0.020)
Observations	299	365	271
Recognized	0.028	0.043	-0.002
	(0.025)	(0.032)	(0.007)
Observations	4283	3011	2124
Exemplary	-0.028	-0.028	0.007
	(0.032)	(0.032)	(0.008)
Observations	1944	2597	2182

Notes: Standard errors clustered by running variable in parentheses. Samples are restricted to plus or minus ten percentage points from each threshold. *** p<0.01, ** p<0.05, * p<0.1

Table 7: RDD Intent to Treat Estimates of Rating Effects on the Probability of Labor Market Success, Plus or Minus Ten Percentage Points Bandwith

	Composite Success	Within-District Success	Salary Rate of Change
Acceptable	0.182*	0.292***	0.023
	(0.094)	(0.099)	(0.016)
Observations	762	762	676
Recognized	0.0257	0.038	0.003
	(0.0221)	(0.024)	(0.005)
Observations	5664	5664	5215
Exemplary	0.004	-0.009	0.008
	(0.020)	(0.023)	(0.006)
Observations	4812	4812	4494

Notes: Standard errors clustered by running variable in parentheses. Samples are restricted to plus or minus ten percentage points from each threshold. *** p<0.01, ** p<0.05, * p<0.1

Table 8: RDD Intent to Treat Estimates of Rating Effects on the Probability of Labor Market Success, by Prior Year Unacceptable Rating Status and Bandwidth

	Composite Success	Within-District Success	Salary Rate of Change
Imbens-Kalyanaraman Optimal Bandwidth			
School not rated unacceptable in prior year	0.321**	0.421***	0.047**
	(0.134)	(0.135)	(0.020)
Observations	269	328	252
School rated unacceptable in prior year	-0.458	-0.423	-0.046
	(0.437)	(0.439)	(0.074)
Observations	28	28	33
± *10 Percentage Points Bandwidth*			
School not rated unacceptable in prior year	0.206**	0.320***	0.030*
	(0.102)	(0.108)	(0.016)
Observations	712	712	635
School rated unacceptable in prior year	-0.295	-0.330	-0.045
	(0.345)	(0.347)	(0.068)
Observations	48	48	40

Notes: Standard errors clustered by running variable in parentheses. Samples are restricted to plus or minus ten percentage points from each threshold. *** p<0.01, ** p<0.05, * p<0.1

Table 9: Mean Labor Market Outcomes and Campus Characteristics, by District Size

		District Size		
	All Campuses	Small	Medium	Large
Elementary Campuses in District	**1-204**	**1-7**	**8-29**	**30-204**
Outcomes: Two Years After Rating				
Within-District Success	79.42	71.83	81.39	84.43
Outside-of-District Success	4.57	8.12	3.73	2.13
Not Success Within District	5.75	5.39	5.86	5.97
Not Success Outside District	2.40	3.69	2.00	1.61
Leave Texas Public Schools	7.87	10.97	7.01	5.86
Campus Student Demographics Shares				
Disadvantaged	0.59	0.54	0.56	0.67
Black	0.13	0.10	0.13	0.18
Hispanic	0.46	0.32	0.49	0.56
White	0.38	0.57	0.35	0.23
Asian	0.03	0.01	0.04	0.04
Special Education	0.11	0.12	0.10	0.10
Gifted and Talented	0.06	0.06	0.06	0.06
Limited English Proficient	0.21	0.10	0.22	0.29
English Second Language	0.06	0.05	0.05	0.07
Observations	11,428	3,582	3,992	3,854

Notes: Sample includes all principals that had at least 2 years of tenure in the school and 25 or fewer years of total experience in Texas public schools in school years 2000-01 to 2007-08 (excluding 2002-03). Two-year success is defined as retaining current principal position, moving to a new position with above median change in salary, or moving to a new principal position with above median change in student achievement.

Table 10: Estimated OLS Effects of School Performance Metrics on the Probability of Labor Market Success, by District Size

	Small Districts		Medium Districts		Large Districts	
	Composite Success	**Salary Rate of Change**	**Composite Success**	**Salary Rate of Change**	**Composite Success**	**Salary Rate of Change**
Pass Rate	0.598***	0.061	0.702***	0.130***	0.249	0.096**
	(0.178)	(0.044)	(0.159)	(0.036)	(0.149)	(0.044)
Change in Pass Rate	-0.203	-0.023	-0.024	-0.055*	0.039	-0.006
	(0.167)	(0.047)	(0.129)	(0.032)	(0.102)	(0.025)
Value-added	0.048	0.021**	0.008	-0.010	-0.029	-0.004
	(0.042)	(0.010)	(0.045)	(0.013)	(0.048)	(0.009)
Accountability Rating						
Unacceptable	-0.259***	-0.041*	-0.176*	-0.021	-0.269***	-0.025
	(0.082)	(0.023)	(0.091)	(0.014)	(0.032)	(0.017)
Recognized	0.038*	0.002	0.030*	0.003	0.004	0.004
	(0.019)	(0.005)	(0.017)	(0.003)	(0.012)	(0.004)
Exemplary	0.042	0.000	0.048*	0.010*	0.002	0.013*
	(0.028)	(0.007)	(0.026)	(0.006)	(0.026)	(0.007)
Observations	3,582	3,189	3,992	3,712	3,854	3,628
R-squared	0.039	0.050	0.036	0.056	0.028	0.085

Notes: Robust standard errors in parentheses. Acceptable is the excluded rating category. All specifications include year fixed effects and the following controls: indicators for principal ethnicity, degree level, gender, tenure, and proportions of students who are black, Hispanic, white, Asian, economically disadvantaged, classified as ESL, classified as gifted and talented, and classified as special education. *** p<0.01, ** p<0.05, * p<0.1

Table 11: Estimated Multinomial Logit Effects of School Performance Metrics on the Probability of Labor Market Success Within District and Out of District, by District Size

	District Size		
	Small	**Medium**	**Large**
Within-District Success			
Pass Rate	3.415***	5.493***	2.021*
	(1.071)	(1.177)	(1.218)
	[0.452]	[0.662]	[0.236]
Change in Pass Rate	-1.580	-0.266	0.125
	(1.003)	(0.995)	(0.748)
	[-0.315]	[-0.013]	[-0.020]
Value-added	0.272	0.057	-0.239
	(0.268)	(0.337)	(0.419)
	[0.033]	[-0.005]	[-0.018]
Accountability Rating			
Unacceptable	-1.260***	-0.694	-1.486***
	(0.391)	(0.487)	(0.165)
	[-0.256]	[-0.125]	[-0.283]
Recognized	0.268**	0.267**	0.044
	(0.118)	(0.128)	(0.106)
	[0.046]	[0.050]	[-0.001]
Exemplary	0.396**	0.457**	0.048
	(0.192)	(0.212)	(0.248)
	[0.068]	[0.063]	[-0.000]
Out-of-District Success			
Pass Rate	3.399**	4.661**	1.226
	(1.616)	(1.958)	(1.924)
	[0.049]	[0.003]	[-0.010]
Change in Pass Rate	0.598	-0.871	2.015
	(1.542)	(2.356)	(2.488)
	[0.115]	[-0.023]	[0.039]
Value-added	0.339	0.459	-0.684
	(0.418)	(0.711)	(0.659)
	[0.008]	[0.020]	[-0.010]
Accountability Rating			
Unacceptable	-0.322	0.0911	0.061
	(0.591)	(0.823)	(0.579)
	[0.039]	[0.029]	[0.043]
Recognized	0.040	-0.393*	0.352
	(0.197)	(0.219)	(0.337)
	[-0.011]	[-0.021]	[0.007]
Exemplary	-0.017	-0.089	0.323
	(0.306)	(0.391)	(0.636)
	[-0.019]	[-0.015]	[0.006]
Observations	3,582	3,992	3,854

Notes: Robust standard errors in parentheses. Average marginal effects are reported in brackets for pass rate, change in pass rate and value-added. Differences in probability of outcomes are reported in brackets for accountability ratings. Acceptable is the excluded rating category. All specifications include year fixed effects and the following controls: indicators for principal ethnicity, degree level, gender, tenure, and proportions of students who are black, Hispanic, white, Asian, economically disadvantaged, classified as ESL, classified as gifted and talented, and classified as special education. *** p<0.01, ** p<0.05, * p<0.1

Table 12: RDD Intent to Treat Estimates of Rating Effects on the Probability of Labor Market Success, by District Size and Bandwidth

	Small Districts			Medium Districts			Large Districts		
	Composite Success	Within-District Success	Salary Rate of Change	Composite Success	Within-District Success	Salary Rate of Change	Composite Success	Within-District Success	Salary Rate of Change
Imbens-Kalyanaraman Optimal Bandwidth									
Acceptable	0.395	0.312	0.019	-0.146	0.0419	-0.032	0.296	0.408*	0.085***
	(0.266)	(0.227)	(0.037)	(0.209)	(0.265)	(0.049)	(0.211)	(0.217)	(0.029)
Observations	103	178	131	146	101	45	182	177	103
Recognized	-0.011	-0.023	-0.003	0.024	0.056	-0.006	0.034	0.055	0.012
	(0.064)	(0.053)	(0.013)	(0.059)	(0.040)	(0.012)	(0.050)	(0.042)	(0.008)
Observations	930	828	810	922	1315	725	801	745	1536
Exemplary	0.004	-0.024	-0.014	-0.063	-0.088	0.023*	-0.018	-0.006	0.004
	(0.039)	(0.055)	(0.014)	(0.053)	(0.058)	(0.012)	(0.045)	(0.047)	(0.012)
Observations	1329	1044	1038	699	691	1141	804	862	626
± 10 Percentage Points Bandwidth									
Acceptable	0.197	0.304	0.006	-0.131	0.073	-0.029	0.270*	0.331**	0.051**
	(0.201)	(0.207)	(0.031)	(0.202)	(0.229)	(0.037)	(0.162)	(0.167)	(0.024)
Observations	233	233	195	180	180	160	349	349	321
Recognized	0.025	0.012	0.002	0.045	0.042	-0.003	-0.005	0.023	0.012
	(0.043)	(0.037)	(0.009)	(0.039)	(0.033)	(0.008)	(0.034)	(0.025)	(0.008)
Observations	1846	1846	1652	1967	1967	1824	1851	1851	1739
Exemplary	0.002	-0.022	-0.008	0.023	0.009	0.019*	-0.016	-0.011	0.011
	(0.036)	(0.044)	(0.011)	(0.032)	(0.036)	(0.010)	(0.035)	(0.037)	(0.009)
Observations	1519	1519	1384	1832	1832	1724	1461	1461	1386

Notes: Standard errors clustered by running variable in parentheses. Samples are restricted to plus or minus ten percentage points from each threshold. *** p<0.01, ** p<0.05, * p<0.1

Appendix Table A1: RDD First-stage Ratings Coefficients for Labor Market Success, by District Size and Bandwidth

	All Campuses			Small Districts			Medium Districts			Large Districts		
	Composite Success	**Within-District Success**	**Salary Rate of Change**	**Composite Success**	**Within-District Success**	**Salary Rate of Change**	**Composite Success**	**Within-District Success**	**Salary Rate of Change**	**Composite Success**	**Within-District Success**	**Salary Rate of Change**
Imbens-Kalyanaraman Optimal Bandwidth												
Acceptable	0.831***	0.841***	0.816***	1.014***	1.038***	1.047***	0.616***	0.592***	0.479	0.853***	0.852***	0.835***
	(0.062)	(0.058)	(0.068)	(0.017)	(0.038)	(0.044)	(0.187)	(0.211)	(0.325)	(0.103)	(0.105)	(0.138)
Observations	299	365	271	103	178	131	146	101	45	182	177	103
Recognized	0.976***	0.973***	0.965***	0.964***	0.967***	0.958***	0.958***	0.965***	0.949***	0.994***	0.993***	0.990***
	(0.008)	(0.011)	(0.014)	(0.022)	(0.019)	(0.024)	(0.022)	(0.018)	(0.029)	(0.009)	(0.009)	(0.007)
Observations	4283	3011	2124	930	1143	810	922	1183	725	801	725	1536
Exemplary	0.939***	0.948***	0.941***	0.964***	0.963***	0.959***	0.910***	0.909***	0.933***	0.964***	0.965***	0.956***
	(0.016)	(0.013)	(0.014)	(0.013)	(0.016)	(0.014)	(0.031)	(0.032)	(0.020)	(0.020)	(0.020)	(0.024)
Observations	1944	2597	2182	1329	1044	1038	699	691	1141	804	862	626
± 10 Percentage Points Bandwidth												
Acceptable	0.881***	0.881***	0.863***	1.022***	1.022***	1.028***	0.622***	0.622***	0.572**	0.853***	0.853***	0.887***
	(0.045)	(0.045)	(0.052)	(0.022)	(0.022)	(0.027)	(0.182)	(0.182)	(0.228)	(0.078)	(0.078)	(0.086)
Observations	762	762	676	233	233	195	180	180	160	349	349	**321**
Recognized	0.978***	0.978***	0.976***	0.966***	0.966***	0.964***	0.977***	0.977***	0.975***	0.991***	0.991***	0.990***
	(0.007)	(0.007)	(0.007)	(0.013)	(0.013)	(0.015)	(0.013)	(0.013)	(0.014)	(0.006)	(0.006)	(0.006)
Observations	5664	5664	5215	1846	1846	1652	1967	1967	1824	1851	1851	1739
Exemplary	0.965***	0.965***	0.962***	0.965***	0.965***	0.961***	0.956***	0.956***	0.953***	0.976***	0.976***	0.975***
	(0.007)	(0.007)	(0.008)	(0.012)	(0.012)	(0.013)	(0.014)	(0.014)	(0.014)	(0.011)	(0.011)	(0.012)
Observations	4812	4812	4494	1519	1519	1384	1832	1832	1724	1461	1461	1386

Notes: Standard errors clustered by running variable in parentheses. Samples are restricted to plus or minus ten percentage points from each threshold. *** p<0.01, ** p<0.05, * p<0.1

Appendix Table A2: Tests of Balance of Student and Principal Characteristics by Rating Threshold

	Acceptable	Recognized	Exemplary
Student Demographic Shares			
Black	0.029	0.001	0.006
	(0.047)	(0.009)	(0.008)
White	0.053	-0.015	-0.038*
	(0.042)	(0.015)	(0.017)
Hispanic	-0.082	0.012	0.032
	(0.057)	(0.017)	(0.017)
Econ. Disadvantaged	-0.027	0.010	0.040**
	(0.033)	(0.013)	(0.015)
Gifted and Talented	-0.006	-0.005*	0.001
	(0.008)	(0.003)	(0.003)
Special Education	0.013*	-0.000	0.004
	(0.006)	(0.001)	(0.002)
Principal Characteristics			
Black	-0.011	0.005	0.019
	(0.082)	(0.016)	(0.013)
Hispanic	-0.186*	-0.008	0.018
	(0.087)	(0.023)	(0.0201)
White	0.198*	-0.001	-0.030
	(0.094)	(0.025)	(0.023)
Years Tenure	0.166	-0.022	0.300*
	(0.368)	(0.119)	(0.128)
Years Experience	-1.403	0.237	0.469
	(1.053)	(0.274)	(0.291)
Bachelor	0.045	0.005	0.004
	(0.052)	(0.013)	(0.013)
Master	0.029	0.009	-0.000
	(0.063)	(0.016)	(0.017)
Doctorate	-0.074*	-0.005	-0.008
	(0.038)	(0.009)	(0.010)
Male	0.096	-0.037	-0.016
	(0.087)	(0.024)	(0.025)
N	762	5665	4812
F-stat	1.085	0.910	1.771
p-value	0.364	0.552	0.0325

Notes: Standard errors in parentheses. Each cell reports the coefficient from a separate seemingly unrelated regression of the first stage equation where the outcomes have been replaced by student shares and principal characteristics. F-statistics and their p-values are reported for the test with a null hypothesis that all coefficents in the column are jointly equal to zero. *** p<0.01, ** p<0.05, * p<0.1