SCHOOL ACCOUNTABILITY RATING SHOCKS,
ACHIEVEMENT, AND SCHOOL CHOICE

Patrick L. Baude
University of Illinois at Chicago
pbaude2@uic.edu

September 2014

**Preliminary and Incomplete**
**Please do not cite without permission**

## Abstract

*By providing information about school quality, state and federal accountability systems are meant to help families make informed choices between schools, and to aid administrators in understanding where effective education practices are being implemented and where improvement is needed. This paper uses student level panel data from Texas from 1998-2011 to estimate how rating differences between schools affect the family school choice decision and subsequent school achievement. I use a fuzzy regression discontinuity design to estimate the causal effects of a rating shock on these school and family responses. I find that a negative shock to the lowest rating level results in an increase in the math and reading pass rates the following year, while shocks at higher rating levels result in little and possibly negative effects on future test scores. Reduced form school choice results suggest that a higher likelihood of a negative rating shock results in small reductions in the likelihood of reenrollment at a traditional public school and larger reductions at charters. In auxiliary school fixed effects analyses, I find that the difference between the highest and lowest ratings is associated with large differences in subsequent school pass rates and in the mean prior achievement of newly enrolled students.*

JEL Codes: H0, H75, I20, I28

This paper represents preliminary work towards my doctoral dissertation in economics at the University of Illinois at Chicago and was prepared for presentation at CERGE-EI, September 16, 2014. The conclusions of the research do not necessarily reflect the opinion or official position of the THECB, the TEA or the State of Texas.

## 1. Introduction

School accountability and school choice have been hallmarks of the standards-based educational reform movement that saw the passage of the *Improving America's Schools Act* of 1994 and the *No Child Left Behind* legislation of 2001. Proponents of accountability and choice hope that by introducing market-based competitive forces into the public education sector, these reforms can improve the quality of instruction in America's schools. Two major factors determine the potential strength of these market forces. First is the ability of families to respond to any perceived differences in school quality and second is the tendency for schools to respond to changing patterns of demand and performance. As pointed out by Loeb et al (2011), these demand and supply factors of the education market rely on the availability and reliability of information about school quality. Parental decisions based on incomplete or inaccurate information can result in persistent enrollment at low quality schools while frictions in the dissemination of information about successful innovative practices can hinder the spread of improved learning opportunities.

Using exogenous variation in school accountability ratings from a fuzzy regression discontinuity (RD) design, I estimate the causal effects of new information about school quality. The benefit of this design is that the resulting shocks to ratings represent changes in information about the quality of the affected schools that is orthogonal to true school quality. The procedure in Texas for determining ratings from raw student achievement data introduces two distinct discontinuities in the relationship between achievement and ratings. The first boundary is straightforward and is based on the strict achievement cutoffs that are set for each rating level. The second is more nuanced and results from the greater number of criterion imposed on more diverse schools through the requirements of *No Child Left Behind.* These discontinuities result in marginally different schools facing substantially different probabilities of being treated with a negative rating shock. By comparing outcomes for schools on either side of these discontinuities, I isolate the effect of rating differences that are independent of the underlying school quality.

In this paper, I use this variation to make two contributions to the literature on the role of ratings in the education market. First, I address the role of information about school quality from ratings in the demand for quality schools. When families choose between schools, they make their decision based on all the information available to them regarding school quality. In a full information environment, families would select schools based on their preferences for each dimension of school quality. When information is limited, beliefs about school quality can be influenced from a variety of information sources, such as private social networks, direct observation, media, and published test score or school rating data.

While findings related to the role of the media or social networks are limited and difficult to causally identify (Ball And Vincent (1998)), there are several studies on the effects of published school quality information on school choice. This literature makes two different implicit comparisons when identifying the effect of a rating shock. One is a comparison between families that are exposed to different amounts of information about school quality, and the other is between families that receive information at different times. Survey work in Milwaukee has found that many parents have incomplete or inaccurate information about the schools attended by their children (van Dunk and Dickman (2002)) while experimental work in North Carolina has estimated the effects of having accurate information on school choice patterns by randomly disseminating achievement data to families (Hastings and Weinstein (2008)). Families that receive credible school quality information are more likely to choose high achieving schools than families that do not receive the extra information. A recent paper using data from British Columbia exploits variation over time in the timing of the release of school rating information (Friesen et al (2012)). The authors find that low ratings induce school separation and that low income and non-English speaking families show a lag in the propensity to separate.

Rather than comparing the decisions of families who receive differing amounts of information or who receive information at different times, I compare the decision of families between schools that receive a negative rating shock to schools that do not receive the shock. In this way, I look at differences in the content of information whereas the existing research has focused on differences in the existence of, or in the timing of release, of information. The reduced form school choice estimates in this paper suggest that a higher likelihood of a negative rating shock results in small reductions in the likelihood of reenrollment at a traditional public school, 1-3 percentage points, and larger reductions, up to 10 percentage points, at charter schools. This is evidence that rating shocks that are not indicative of true quality differences do play a role in the family school choice decision and is consistent with the existence of a pure rating effect. In auxiliary school fixed effects analyses, I find that the difference between the highest and lowest ratings is associated with a 0.1-0.2 standard deviation increase in the mean prior achievement level of incoming students, providing evidence that rating can affect the student body profile in ways other than the effect on the likelihood of reenrollment.

Second, I extend the literature on the impact of ratings on future student achievement by estimating the effects in a new policy environment, Texas, whose public schools system educated over 5 million students in 2012.[1] I find that a negative shock to the lowest rating level results in a 4-6% increase in the math pass rate the following year, while shocks at higher rating levels result in little and possibly negative effects on future test scores. This is consistent with a model of education provision where schools respond to the market induced threat of reduced enrollment, and the commensurate funding, when assigned a low rating. Prior findings in this area are mixed. Similar positive effects of a shock to the lowest rating level have been found in Florida (Chiang (2009) and Rouse et al (2013)) and New York City (Rockoff and Turner (2010)) while other work

[1] Source: TEA. This is second in the US behind California.

has found that a failing rating in California results in a decline in future achievement (Sims (2013)).

Just as in the school choice case, I also run school fixed effects regression to estimate the effects of lagged ratings on subsequent school performance. These estimates suggest a 5-9 percentage point difference in subsequent math pass rates when schools receive the highest rating versus the lowest rating further supporting the notion that receipt of a low rating does positively incent schools to improve performance.

In other work dealing with the relationship between school choice and information about school quality, Mizala & Urquiola (2013) find that awards based on national rankings in Chile have been found to have no effect on enrollment, and Hanushek et al (2007) and Baude et al (2014) find that while the charter school reenrollment decision in Texas has been shown to be positively related to school quality as measured by value added or accountability ratings, the mechanism by which families learn about school quality is unclear. In the Florida housing market, Figlio & Lucas (2004) find that the family location decision, which often has a large school choice component, is reflective of the rating assigned to local schools after controlling for other differences between locations that may be correlated with ratings.

This paper proceeds as follows: section 2 briefly motivates the analysis; section 3 describes the accountability system in the state of Texas; section 4 describes the data used in the analysis; section 5 describes the empirical methodology; results for the reenrollment and achievement analyses are in section 6 and 7 respectively; section 8 provides early results regarding incoming student profiles; section 9 addresses threats to validity and section 10 summarizes and concludes with a discussion of policy relevance.

## 2. Motivation and Conceptual Framework [In Progress]

Two of the guiding principles used by the the Texas Education Agency in designing its accountability system are to "increase student performance," and "to enable the public's right to know each school's performance level" and to identify schools in need of reform, and those worthy of recognition.[2] The following conceptual framework regarding the role of information about school quality in the public education market, focuses on two outcomes which are motivated by these guiding principles.

First is the decision made by families about where to send their children to school, or school choice. School accountability ratings are a way of providing information to families about public school quality. Quality is multidimensional, and information about quality comes from many sources besides the accountability rating system. Some dimensions of quality that may matter to families are academics, athletics, peer quality, facilities and civic engagement. When exercising school choice, families may learn about any of these dimensions from a variety of sources. For example, families may learn about schools from personal experience, through private social networks, or via the media in addition to the published accountability rating information. As long as families' utility is increasing in their child's school's mean academic achievement, all else equal, the relative desirability of a school increases in its own rating, and decreases in the ratings of other schools. Therefore the pool of students that would choose to attend one school versus another will differ depending on the quality of each student's outside option. It is important to emphasize that at the cusp between two rating levels, the margin used in the main analysis in this paper, accountability ratings provide little true information about the relative quality of schools. In terms of school choice involving schools on the margin between ratings, if rating information is used by families to select between schools we can estimate the pure effect of information about quality that is independent of other school characteristics including each families' true subjective valuation of school quality.

[2] A full list of the guiding principles behind the Texas accountability rating system is given in Appendix A2.

Second is if schools respond to rating shocks in ways that result in improved standardized test results. If schools are competing for students, and families respond to accountability ratings when choosing schools, then it follows that schools would prefer to be assigned the highest rating possible all else equal in order to avoid loss of enrollment. With funding generally being directly determined by enrollment, the validity of this assumption is strictly a matter of degree. With ratings being primarily determined by campus-level standardized test performance, a school on the cusp between ratings would have motivation to increase achievement so as to ensure the higher rating. However, since there is little statutory incentive associated with being assigned a specific rating[3], if families do not respond to ratings conditional on quality, then schools may be indifferent between ratings and not exhibit differential achievement gains in the subsequent year.

## 3. The Texas School Accountability Rating System

Since 1993, Texas has assigned annual ratings to all public schools through an integrated accountability system. On August 1st every year campuses receive a rating of, in order of decreasing distinction, Exemplary, Recognized, Academically Acceptable, or Academically Unacceptable[4]. Through the parallel Alternative Education Accountability (AEA) system, Alternative Education Campuses (AECs)[5] are rated as either academically acceptable or unacceptable. During the period of analysis in this work, approximately 5% of rated campuses fall under the AEA system. In the remainder of this paper, I focus on schools subject to the standard rating system.

---

[3] See section 3.3 for a discussion of the consequences and rewards associated with each rating level.

[4] Until 2002, the Unacceptable rating was called "Low Performing." Some of the literature using data from Texas during this time period uses this alternate naming convention for the lowest rating.

[5] AECs serve primarily one or more of the following student populations: students at risk of dropping out; recovered dropouts; pregnant or parenting students; adjudicated students; students with severe discipline problems; or expelled students."

From 1991-2002, the standardized exam was the Texas Assessment of Academic Skills (TAAS), which was administered each spring to all students in eligible grades. Mathematics and reading exams were given every year in grades 3-8 and 10 while writing and social studies were only tested in certain grades and years. The standardized exam from 2003-2011 was the Texas Assessment of Knowledge and Skills (TAKS). Several changes accompanied the transition to TAKS: 11th grade evaluations in mathematics and reading were added[6]; campus performance on science exams entered the accountability formula; more difficult exam content was accompanied by a lowering of the passing rate required for each rating; and while tests were administered in 2003, no ratings were assigned in order to ease schools' transition to the new exams.

While the requirements for each rating level have evolved over time, campus pass rates on the TAAS and TAKS exams have been the primary factor determining ratings throughout. In order to achieve a given rating level, campus pass-rates in *all* subjects must meet or exceed the pass rate standard for that rating. Failure to meet the standard for just one subject can reduce a campus' rating irrespective of how well it may have performed in other areas. One dimension of the analysis below uses the pass rate standards as a means of isolating exogenous variation in school ratings.

### 3.1 Subgroup evaluation rules

In addition to the campus wide pass rates on the TAAS and TAKS exams, the system also evaluates schools on the performance of student ethnic and income subgroups. The subgroups eligible for individual evaluation are economically disadvantaged students, whites, blacks, and Hispanics.[7] Campuses where the number of test takers for a given subject belonging to a given

[6] Under TAKS, the 10th and newly added 11th grade reading exams were renamed English Language Arts (ELA). For the purposes of campus-wide subject pass-rates, the accountability system combines the grade 3-8 reading results with the grade 10-11 ELA results. For ease of exposition, I therefore refer to all reading and ELA exams as reading irrespective of grade.

[7] Students are assigned to race-based subgroups by their districts. A student is classified as economically disadvantaged if she comes from a family whose income is below the poverty line,

subgroup exceeds a minimum size requirement (MSR) are evaluated on the performance of that subgroup-subject combination in addition to the campus-wide performance in the subject. The MSR varies between 30 and 50 students by subject and campus. It can be succinctly expressed as $min\left(max\left(30, \frac{examinees}{10}\right), 50\right)$ where *examinees* is the total number of students tested in the subject at the campus. As an example, a school where more than 50 black students sit for the math exam is evaluated on the pass rate among black students in math. In addition, this school is evaluated on the overall pass rate on math for all students including blacks. On the other hand, a school with less than 30 black math test takers is evaluated on the overall pass rate in math but not the black math performance separately. Evaluation for groups between 30 and 50 students depends on the total number of test takers. To illustrate, a school where 400 students sit for the reading exam will be evaluated on the performance of individual subgroups that have 40 or more students taking the reading exam. One dimension of the analysis below uses this discontinuity in group evaluation status across the MSR as a means of isolating exogenous variation in school ratings.

### 3.2 Other Factors Affecting Ratings

In addition to the performance on TAAS/TAKS, at different points in time schools were evaluated on other measures. These include: the annual dropout rate amongst 7$^{th}$ and 8$^{th}$ graders; the annual dropout rate among 7$^{th}$-12$^{th}$ graders; the four-year completion rate amongst 9$^{th}$-12$^{th}$ graders; the State Developed Alternative Assessment (SDAA and SDAA II) for special education students between[8]; and English language learner (ELL) and commended performance rates[9]. In

receives a Pell Grant or funds from a comparable needs-based program, or meets the requirements for one of the following: free or reduced price lunch, Title II of the Job Training Partnership Act, Food Stamps, or Temporary Assistance to Needy Families.

[8] SDAA/SDAA II were in effect from 2004-2007.

[9] ELL and commended performance, introduced in 2011, add further requirements for the Recognized and Exemplary ratings only.

2008 the TAKS-Modified and TAKS-Accommodated exams were incorporated into the TAKS indicators to evaluate students receiving special education services and those with disabilities.

Schools that fail to meet the absolute standards for the next higher rating for every indicator can still move up one rating by meeting certain extra criteria. These criteria are Required Improvement, the Texas Projection Measure, and exceptions provisions. Required Improvement (RI) allows for schools to meet the standards for all indicators if its performance shortfall is less than half of its performance shortfall in the previous year. RI is designed to reward schools who are "on pace" to close the shortfall within two years. RI can be used to move up to Acceptable or Recognized but cannot be used to move up to Exemplary status.

In 2009 and 2010 the TEA used the Texas Projection Measure (TPM) when computing student performance. TPM allows individual student to be counted as meeting the standard for a specific TAKS subject if they were projected to meet the standard for that subject in a future grade (5,7, 8 and 11 are the "to" grades for the projections). The TPM is an out of sample projection using coefficients from a simple OLS regression of past students' future score on their current and past scores, and the average performance at their campus.[10]

After RI and TPM are applied, exceptions provisions can then be used to allow schools to "gate up" to a higher rating. Exceptions can be applied to any of the 26 potential TAKS and ELL indicators, which are within 5 or 10 percentage points of the standard depending on subject and year. Exceptions may not be used for the same subject/subgroup in consecutive years. The number of exemptions increases with the number of indicators that a school is evaluated on as indicated in appendix A3. From 2004 to 2007 exceptions provisions could only be used to move up from Unacceptable to Acceptable. From 2008 onwards exceptions could also be used to move up from Acceptable to Recognized or from recognized to Exemplary. The exceptions act to

---

[10] Details of the calculation of the TPM are beyond the scope of this paper. See http://www.tea.state.tx.us/index3.aspx?id=8351&menu_id=793 for more details.

mitigate the fact that larger and more diverse schools are subject to evaluation on a greater number of indicators.

### 3.3 Consequences and Rewards

The statutory incentives built into the TEA accountability rating system come in the form of both awards and punishments. Schools that receive a low/unacceptable rating are subject to potential disciplinary action and can lose enrollment due to a provision which allows students at low performing schools to transfer to higher performing districts via the Public Education Grant (PEG) program. Consecutive years of low performance can result in school closure. Higher rated campuses are exempted from certain regulations in the Texas Education Code, such as class size requirements.

Between 1992-2006 the Texas Successful Schools Award System (TSSAS) provided funds to schools using a *value added* or *gains* approach. Award funds could only be used for certain items, which excluded teacher compensation and capital projects like gymnasiums. The awards were initially intended to allocate funds to schools that received an acceptable rating or better, which showed high growth in math and reading, and that did not have an "excessive" exemption rate. Nevertheless, in some years unacceptable schools did receive award funds as well. This system had $2.5 million annually to be distributed across all winners, but in starting in 2001 the state legislature did not allocate any funds for the program other than 2002 when it was allocated $500,000. The most that any one school could receive was $5,000 and based on the limited information available schools tended to receive approximately $1,500 when the program was fully funded.

Negative consequences exist only for the unacceptable rating while official rewards are reserved for the exemplary rating. The consequences of a low accountability rating include interventions on the part of the state and federal governments.

The TEA Program Monitoring and Interventions (PMI) Division "develops and implements interventions and sanctions for districts and campuses rated academically unacceptable" and "implements interventions for those campuses that are rated acceptable but that would not satisfy performance standards if the accountability standards for the subsequent year were applied."[11] Exemplary rated campuses gain automatic exemptions from some statutes and rules. Notably, these schools can be exempt from class size requirements. Districts with Unacceptable campuses can be subject to loss of enrollment and the concomitant state/federal funding, suffer the potential loss of accreditation, and are subject to sanctions and the directives of the commissioner.

Since 2007, district accreditation has been assigned based on the district accountability rating and financial accountability rating which in turn are based in large part on the ratings of their member campuses. In this way districts can lose accreditation if they exhibit a pattern of having one or more Unacceptable campuses within the district.

## 4. Data

The longitudinal administrative data for this project consist of restricted access student-level data, and publicly available campus-level data. These data originate from the Texas Education Agency (TEA) and cover the universe of Texas public school campuses and students from 1997-2011.

The publicly available campus-level data can be downloaded from the TEA Academic Excellence Indicator System and Accountability System websites. These datasets include pass rates on standardized tests, campus accountability rating, demographic variables, charter status, enrollment, grades offered, and the number of students in certain demographic subgroups who are tested. Achievement is measured by the performance on statewide standardized tests in mathematics, reading, and in some grades writing, science, and social studies. Figure 1 shows

[11] See the Texas Education Code (TEC), Chapter 39, Subchapter E.

subsequent campus level improvements in pass-rates gains by rating. The mean pass rate gain is monotonically decreasing in the prior year's rating. Reading shows the same pattern.

The restricted access student-level data are accessed via the Texas Schools Project (TSP) at the University of Texas at Dallas.[12] These data follow individual students as they progress through the Texas public schools system allowing for the calculation of reenrollment at the student level and incoming student profiles at the campus level. Key characteristics in this dataset include standardized test scores, grade, demographic variables, attendance, graduation, disciplinary infractions, and characteristics of the academic program attended.

From this student-level panel, it is possible to characterize the dynamics of student transitions along several dimensions. Changes to a school student body profile from one year to the next depend on the relative characteristics of retained students and new students who attended a different school in the prior year. These two factors can be looked at through reenrollment and the incoming student profile.

The relationship between the reenrollment and school ratings is presented in figure 2. Panel A shows the distributions of residual reenrollment rates by school accountability rating after controlling for student ethnicity and low income status, and year dummies. Recognized and acceptable schools have similar distributions while exemplary and unacceptable schools lie to the right and left respectively. The same pattern is evident in panel B where school quality is controlled for using estimates value-added to math test scores.

**5. Empirical Methodology**

Having documented the differences in student mobility patterns and pass rates between schools with different ratings I move on to the estimation of the causal relationship between ratings and these variables. The difficulty in estimating causal effects of school ratings results

[12] A more detailed description of the underlying database can be found in Kain (2001) and other publications on the website for the UTD Texas Schools Project.

from systematic differences between schools with different ratings in terms of student achievement, family characteristics, the availability of outside options, and other unobservable factors. For example, the differences in subsequent achievement and school choice patterns by rating are clearly influenced by school feeder systems and differences in prior achievement and demographic characteristics.

This paper takes two approaches to address these problems to estimate the effects of school ratings on achievement and choice outcomes. The first approach uses a fuzzy regression discontinuity (RD) framework that focuses on schools, which lie on the cusp between rating levels. In this situation it is possible to analyze schools that receive different ratings as though the rating had been randomly assigned. While internally valid, this approach estimates a local average treatment effect (LATE) for a very specific sets of schools based on a very specific source of rating variation. The second approach takes advantage of the panel nature of the data by estimating school fixed effects (FE) models. Estimates from this approach are identified by variation in outcomes over time within the subset of schools that receive at least two different ratings during the sample period. As opposed to the RD strategy, any school that experiences a rating change from one year to the next at any time contributes to the identification of the FE estimates. Drawbacks of the FE strategy relate to the implicit assumptions regarding the persistence of student performance and the role of student and campus unobservables. The remainder of this section describes the details of these two approaches to identification.

### 5.1 Fuzzy Regression Discontinuity

The implicit comparison in the fuzzy RD design is between schools that have large differences in the probability of receiving a lower rating due to marginal differences in academic performance or due to marginal differences in demographic composition. Due to the ordinal nature of the four rating levels in the Texas system, the effect of these negative rating shocks need to be assessed independently at each margin. For ease of exposition, I limit the rating space to two levels, acceptable and unacceptable without loss of generality. This approach then can be applied separately at each of the three rating margins, *Exemplary-Recognized*, *Recognized-Acceptable* and *Acceptable-Unacceptable*.

The discontinuity related to academic achievement can not be combined with the discontinuity related to demographics into a single running variable due to fundamentally different units of measure between variables.[13] This fact necessitates the use of a multidimensional RD design. Reardon & Robinson (2012) outline ways to implement a regression discontinuity design with multiple running variables. To identify the effect of an unacceptable rating, I use a combination of what they call a *binding score RD* and either the *fuzzy frontier RD* or the *surface RD*. The binding score RD is well suited to the all-or-none aspect of the rating system whereby failure to meet just one criteria can be sufficient for a downgrade, and the frontier RD component is a direct consequence of having to analyze the discontinuity at the achievement-based boundary separately from the demographics-based boundary.

In the following detailed description of how I construct the running variables, I refer to subgroup-subject combinations, indicators, that have a pass rate below the requirement for a particular rating as a 'low performing indicator' and since not all subgroup-subject combination pass-rates enter into every campus' rating formula, I refer to each indicator that is used as an 'evaluated indicator.[14] Recall from section 3 that an indicator is only evaluated if the number of students in the indicator (e.g. the number of low income students tested in math) exceeds the campus-subject specific minimum size requirement (MSR) for that indicator. A school can be dropped down a rating level if just one evaluated indicator is low performing. This results in an ex-post campus-level probability of receiving a negative rating shock that depends discontinuously on both the size and the performance of each subject-subgroup combination.

---

[13] Pass-rates are measured in terms of percentage points, and student group sizes are measured by numbers of students.

[14] Appendix A1 lists the pass rates that a campus must achieve for all evaluated subgroup-subjects to achieve each rating.

Treatment with a higher probability of receiving a lower rating can occur two ways. First, conditional on being evaluated, an indicator induces treatment when it crosses the boundary from high to low performing. Second, conditional on being low performing, an indicator induces treatment when it includes enough students that it crosses the MSR and becomes evaluated. These discontinuities produce two marginal indicators for each campus: the lowest performing evaluated indicator and the largest underperforming indicator. The position of these marginal indicators relative to the MSR and the required pass rate determine whether or not a campus is treated and whether or not it falls within the specified bandwidth for inclusion.

Since the marginal indicator is almost always not the same when evaluating the pass-rate and MSR induced discontinuities, I refer to these indicators as the 'marginal pass-rate indicator' and 'marginal MSR indicator' respectively. Schools whose marginal pass-rate indicator falls just below the pass rate standard face a higher ex-post probability of being treated with a rating shock when compared to schools whose marginal pass-rate indicator just meets the same pass rate standard. Within a small percentage point bandwidth of the standard being above or below is essentially random, and any differences in outcomes can be interpreted as having been caused by the differential treatment probability.

Schools whose marginal MSR indicator falls just above the MSR face a higher probability of being treated with a rating shock when compared to schools whose marginal MSR indicator falls just short of meeting the MSR. Similar to the first discontinuity, this comparison yields the causal effect of a rating shock under the assumption that being to the right or left of the MSR is random conditional on being within a small number of student bandwidth of the MSR. Kane & Staiger (2002) discuss this discontinuity as it relates to the fairness of school accountability systems.

Figure 3 shows the interrelatedness of the two dimensions of treatment graphically. The relevance of each indicator depends on where it lies in the 'pass rate' – 'subgroup size' space. For generality, the pass rate and subgroup size have been centered to be zero at the required pass rate and minimum size requirement cut-offs for each subject-campus-year. Negative values for the centered pass rate on the y-axis and positive values for the centered subgroup size on the x-axis are needed for treatment. Small changes in the size or performance of the marginal indicators can result in a campus being moved into or out of treatment. A campus is treated if for *any one* of the 20 indicators *both* of the following conditions hold: the pass rate for the indicator is below the standard for that rating for that subject in that year; and the number of students in the indicator meets or exceeds the campus MSR for that subject that year. If either of these conditions is not met, then the indicator does not induce treatment.

Indicators that fall in quadrant I are large enough to be evaluated and exhibit a high enough pass rate to qualify for the higher rating. Indicators that fall in quadrant II exhibit a pass rate that exceeds that required for the high rating, but are too small to be evaluated. Indicators in quadrant III have a pass rate too low for the higher rating, but are too small for this shortcoming to induce a rating shock. Lastly, indicators in quadrant IV have a low pass rate and are large enough to be evaluated. Any campus with an indicator in this last quadrant is more likely to receive a rating shock than those with no indicator in this quadrant.

In the context of figure 3, the regression discontinuity design I use follows a combination of two of the approaches defined in Reardon and Robinson's (2012) summary of methods for implementing regression discontinuity designs where multiple factors affect treatment. These are what they term the *binding score regression discontinuity* and the *fuzzy frontier regression discontinuity*. The binding score aspect involves the reduction of all indicators into marginal MSR and pass-rate indicators. The frontier component results in two separate sets of estimates, one for schools near the treatment boundary between quadrants I-IV and another for schools on the treatment boundary between quadrants III-IV. The I-IV frontier corresponds to being moved into treatment by having group that is a low performing group conditional on being evaluated. The III-

IV frontier corresponds to being moved into treatment by having a single group become evaluated conditional on it being low performing. Table 1 summarizes these two dimensions of treatment.

Panels (a) and (b) of figure 4 illustrate specific examples of campuses that are on either side of the pass-rate frontier (I-IV). The campus depicted in panel (a) has three indicators that are large enough to be evaluated. These are Hispanic students in math (*hm*), Hispanic student in reading (*hr*) and economically disadvantaged students in math (*em*). The *em* indicator functions as the marginal pass-rate indicator since it has the lowest pass rate relative to the standard. Since the pass rate for the marginal indicator is below the standard, this campus is *treated* with a higher probability of receiving a shock from recognized to acceptable. The pass rate for the marginal pass-rate indicator (*hm)* for the campus depicted in panel (b) on the other hand exceeds the standard required for a recognized rating and therefore this campus is *untreated*. Panels (c) and (d) illustrate a specific example of two campuses that are on either side of the MSR frontier (III-IV). The campus depicted in panel (c) has five indicators that are low performing enough to induce a rating downgrade. These are Hispanic students in science (*hc*), economically disadvantaged students in writing (*ew*), white students in writing (*ww*), black students in science (*bc*), and economically disadvantaged students in science (*ec*). The *ec* indicator functions as the marginal MSR indicator since it is the largest relative to its MSR. Since the size of the binding MSR indicator exceeds the threshold, this campus is *treated* with a higher probability of receiving a negative rating shock from recognized to acceptable. The size of the binding MSR indicator (*hc*) for the campus depicted in panel (b) on the other hand is too small to induce evaluation and therefore this campus is *untreated*.

In terms of implementation, I analyze this fuzzy RD design based on the following instrumental variables regression:

$$y_{it} = \theta^{IV} U_{it-1} + f(D_{it-1}) + X_{it}\beta + \tau_t + \varepsilon_{it}, \tag{1a}$$

where $y_{it}$ is the outcome for school $i$ in year $t$.[15] This is regressed against a dummy indicating if school $i$ receives an unacceptable rating in the prior year, $U_{it-1}$, and flexible function $f(\cdot)$ of the running variable $D_{it}$. $X_{it}$ is a vector of student and campus demographics, and $\tau_t$ represents a set of year dummies. While not strictly necessary, the inclusion of covariates, $X$, improves precision especially when bandwidth restrictions reduce the sample size.

To instrument for the endogenous rating variable $U_{it}$, I use the treatment dummy, $T_{it-1}$ which is equal to 1 if the school's binding indicator falls in quadrant IV in figure 3 and 0 otherwise. This can be decomposed into the reduced form:

$$y_{it} = \theta^{RF} T_{it-1} + f(D_{it-1}) + X_{it}\beta + \tau_t + \eta_{it}, \tag{1b}$$

and first stage:

$$U_{it-1} = \theta^{FS} T_{it-1} + f(D_{it-1}) + X_{it}\beta + \tau_t + \zeta_{it}. \tag{1c}$$

The reduced form coefficient $\theta^{RF}$ gives the effect of having an evaluated indicator that is low performing regardless of if that indicator results in an unacceptable rating. The first stage coefficient $\theta^{FS}$ measures the increase in the probability of receiving an unacceptable rating when a campus is treated with an evaluated indicator that is low performing.

For the running variable control function $f(\cdot)$ I use a non-parametric local linear regression (LLR). The treatment dummy is defined to be one if the campus' binding indicator falls The identifying assumption that $T_{it}$ only affects $y_{it}$ through its effect on $U_{it}$ is achieved by limiting the sample to schools which lie within a small bandwidth of the discontinuity and their students.

[15] Alternatively, the index $i$ can refer to students. In which case andy school-level variables indexed by $i$ will refer to the school attended by student $i$.

Concerns about the ability of schools to differentially sort out of treatment is mitigated by the fact that the final determination of treatment is only realized on on the testing day when the number of test-takers is determined and the exams are taken. At the beginning of the school year, schools may have knowledge about the size and past performance of students in each indicator leading to the possibility that schools may attempt to alter the ex-ante probabilities of different rating outcomes. Nevertheless, by restricting the analysis to schools near the boundary and the random components of attendance and test scores it is difficult for schools to precisely manipulate their ex-post probability of treatment. Nevertheless, this assumption could be directly tested if the initial school rosters at the start of the year could be compared to the list of students who end up sitting for the standardized tests later in the year. As these data are recorded for certain years, this is a potential avenue for further research.

**5.1.1 First Stage**

Each specific outcome variable and analysis in the results sections has its own specific first-stage F-statistic due to differences in samples. For that reason, I will present the first stage effects of the marginal indicator pass-rates and sizes on the probability of a negative rating shock graphically using the entire sample. Figure 5 illustrates the first-stage effects of having a marginal pass-rate indicator that lies on one side of the pass-rate requirement or the other. In these figures, I plot the proportion of schools rated below the specified rating level on the vertical axis against the number of percentage points away from the pass-rate requirement for each campus' marginal pass-rate indicator on the horizontal. In this case treatment occurs to the left of the discontinuity. Panel (c) for example gives the proportion of schools rated recognized or below by the distance in percentage points from the exemplary pass-rate requirement for the marginal pass-rate indicator at the campus. For schools whose marginal pass-rate indicator just meets the required pass-rate for exemplary (meaning that the lowest performing group that is evaluated meets the exemplary cut-off) the proportion of schools rated below exemplary drops approximately 50 percentage points. Similarly in panel (b) we can see that the proportion of schools that miss out on the recognized rating drops almost 40 percentage points once the marginal pass-rate group meets the standard for the recognized rating. One specific aspect of the system is clearly visible in panel (b). The *Required Improvement* provision allows for schools that should be rated acceptable but are within five percentage points of the recognized standard to be moved up to the recognized rating if they have shown evidence of past performance gains that would lead to meeting the standard within two years. For this reason the probability of being rated below recognized exhibits two discontinuities, one at the standard, and a smaller one at 5 points below the standard. This additional discontinuity does not affect results, as the maximum bandwidth used is 5 percentage points.

Note that the discontinuity in panel (a) is much smaller, less than five percentage points. This is due to the other aspects of the accountability system outlined in section 3.2, notably the exceptions provisions, which help schools that would have otherwise been assigned an unacceptable rating to gain a bump up to the acceptable rating. For this reason, the instrument at this boundary is weak leading me to omit some of the IV estimates of a rating shock to unacceptable based on the pass-rate discontinuity from the tables in favor of the reduced form.

Figure 6 illustrates the first-stage effects of having a marginal MSR indicator that lies on one side of the MSR or the other. In these figures, I plot the proportion of schools rated below the specified rating level on the vertical axis against the number of students away from the pass-rate requirement for each campus' marginal MSR indicator on the horizontal. In this case treatment occurs to the right of the discontinuity. Panel (b) for example gives the proportion of schools rated acceptable or below by the distance in terms of the number of students from the MSR for the marginal MSR indicator at that campus. For schools whose marginal MSR indicator just meets the MSR (the largest low performing group is evaluated) the proportion of schools rated below recognized jumps approximately 50 percentage points. Similarly in panel (c) we can see that the proportion of schools that miss out on the exemplary rating increases approximately 45 percentage points once the marginal MSR group meets the MSR.

In all, with the exception of the pass-rate induced discontinuity at the *Acceptable-Unacceptable* boundary, the first stage effects of treatment on the likelihood of receiving a lower rating are large, and statistically significant.

### *5.2 School Fixed Effects*

While the RD approach estimates separate effects for the rating differential at each rating boundary, the school fixed-effects (FE) strategy compares the outcomes from each rating against an excluded group in a single regression. This approach identifies the effects of receiving a certain rating based on variation within a campus over time in the assigned rating and in achievement and enrollment patterns.

In the school choice portion of the analysis I implement the FE regression on a dependent variable measured in levels:

$$y_{it} = \sum_{A=\{E,R,U\}} \theta^A A_{it-1} + X_{it}\beta + \tau_t + \gamma_i + \varepsilon_{it}, \tag{1}$$

where $y_{it}$ in the student level regressions is a dummy equal to one if student $i$ reenrolls at the same school in year $t$ and zero otherwise. In the campus level regressions the outcome is defined as various mean characteristics of all students in their first year at school $i$ in year $t$. The $\theta$ coefficients on the lagged accountability rating dummies give the mean difference in the probability of reenrollment and incoming student profiles for schools, or students at schools, who received exemplary, recognized and unacceptable ratings in the prior year relative to the excluded group of acceptable schools. I control for school specific time varying characteristics $X$ in addition to the school and year dummies, $\gamma_i$ and $\tau_y$.

When evaluating the role of ratings on the next year's performance in a campus FE regression, I estimate two different specifications. First is a lagged dependent variable model:

$$PR_{it} = \lambda PR_{it-1} + \sum_{A=\{E,R,U\}} \theta^{A} A_{it-1} + X_{it}\beta + \tau_t + \gamma_i + \varepsilon_{it}, \tag{2}$$

where $PR_{it}$ is the pass rate for school $i$ in year $t$. The $\theta$ coefficients on the lagged accountability rating dummies give the mean difference in the pass rate for schools who received exemplary, recognized and unacceptable ratings in the prior year relative to the excluded group of acceptable schools.

As pointed out by Wooldridge (2002), this lagged dependent variable model violates the conditional strict exogeneity assumption that the independent variables are uncorrelated with not only the contemporaneous error, but also the errors in all other time periods. Here, this is violated because by construction $cov(PR_{it-1}, \varepsilon_{it}) \neq 0$ for all $t$. I partially address this by instrumenting for the lagged pass rate with the twice-lagged pass rate.[16] As Imberman (2011) points out, the resulting estimate in conjunction with the results from the gains model below can be used to provide bounds for the true rating effect.

As alluded to above, I also estimate a gains model that is a special case of equation (2) where the persistence parameter $\lambda$ from the lagged dependent variable model is restricted to be one:

$$\Delta PR_{it} = \sum_{A=\{E,R,U\}} \theta^{A} A_{it-1} + X_{it}\beta + \tau_t + \gamma_i + \varepsilon_{it}. \tag{3}$$

In all likelihood, the persistence of the campus pass-rate likely takes on a value less than one rendering the gains model invalid. However, following Imberman (2011), in conjunction with estimates from equation (2) estimates from equation (3) can be used to form bounds on the true value of the role of the lagged rating on the current pass-rate.

[16] A similar strategy is implemented in estimating teacher value-added in Jacob, Lefgren and Sims (2002) and in estimating charter school quality in Hanushek et al (2007). Later work by Todd and Wolpin (2007) however, brings up important questions regarding the exogeneity of the twice lagged pass rate.

## 6. Reenrollment Results

In this section I give all results separately by school type, charter or TPS. As a consequence of the structure of the education system, charter enrollees have a differential ability to quickly respond to a ratings release than students attending a TPS. Charter school students always have the outside option of attending their local TPS. The same cannot be said for TPS students who may have an outside charter option that is unable to accept more students. Empirically, charter school students in Texas exhibit greater mobility than traditional public school students (Hanushek (2007)). A portion of this difference may be indicative of differences in the characteristics of schools across sectors or differential sensitivities to relative school quality for the families of charter enrollees versus traditional public school students. Furthermore, since charter students by definition have all exercised school choice in entering a charter to begin with, charter enrollees may exhibit a greater sensitivity to measures of school quality relative to their TPS peers. For all of these reasons, estimates are given for charter and TPS students separately. Additionally, at times results are further broken down by one- and two-year ahead response.[17]

### 6.1 Regression Discontinuity

Table 2 presents the reenrollment results for students enrolled at traditional public schools (TPS) within a 3 percentage point or student bandwidth. In panel (A) we can see that at all rating thresholds for both instruments, the negative rating shock has a small or negative effect on likelihood of reenrollment next year. Panel (B) highlights the short response horizon between the release of ratings and the start of the next year. Here the two-year response is larger in magnitude for all rating-boundaries for both instruments. While the estimates vary by rating boundary and instrument, the 2-year ahead reduced form estimates indicate that the probability of reenrollment drops 1 to 3 percentage points when a school is treated with a low performing

[17] Given the late July or early August rating release schedule, the one-year ahead reenrollment figures correspond to a 1-2 month response window, while the two-year ahead figures correspond to a 13-14 month response window.

evaluated indicator. With the incorporation of the first stage effects, the IV estimates indicate up to a 10 percentage point drop in the 2-year ahead reenrollment probability when a school suffers a rating shock. With the base 2-year reenrollment rate at approximately 50% this decrease represents a large 20% drop in the reenrollment likelihood.

Consistent with the notion that charter enrollees are more mobile than their TPS counterparts, table 3 shows the much larger reenrollment response among charter students to accountability rating shocks. Reduced form estimates report a drop in the reenrollment rate of between 4-18 percentage points. I omit the IV estimates due to concerns regarding the first stage effects. Preliminary IV estimates that are excluded from the table show implausibly large effects resulting from very weak first stages. I plan to address this issue in later drafts of the paper. Nevertheless, as upwards of 97% of the students in the sample are enrolled at a TPS the results for TPS alone clearly demonstrate the adverse reenrollment effects of a negative rating shock.

**6.2 School Fixed Effects**

Table 4 gives results of the student level reenrollment regressions. Here we see the same qualitative pattern as in the raw data where schools with higher ratings exhibit greater reenrollment rates. This is true for both charter and TPS, and for both one- and two-year ahead reenrollment. Charter students are more sensitive one year ahead than two years ahead, while TPS are more sensitive 2yrs ahead. This is consistent with the notion that TPS students must either move or win entry into a charter in order to switch schools, whereas charter students need not move nor go through the application process to switch back to the local TPS.

Even numbered specifications include campus fixed effects limiting variation to changes within schools over time in the rating assigned and the subsequent reenrollment probabilities. Almost all charter coefficients lose significance and drop in magnitude. With TPS however, the one-year response is relatively similar with the exception of the effect of an unacceptable rating, which drops by a factor of six. Similar patterns emerge when using the 2-year ahead reenrollment probability. In all, after the inclusion of school fixed effects, there is very little difference in reenrollment rates that can attribute to schools being assigned different accountability ratings. Using estimates from Hanushek, Kain and Rivkin (2004) regarding the effects of student mobility on overall campus achievement suggests that the maximum 1.3 percentage point change in one-year ahead reenrollment would translate to a very small 0.002 standard deviation change in test scores.[18]

## 7. School Performance Results: Standardized Test Pass Rate

### 7.1 Regression Discontinuity

Table 5 summarizes estimates of the effect of a negative rating shock on the next year's math pass rate. Panel (A) uses the position of the pass-rate binding indicator relative to the threshold as an instrument for the school rating. The resulting identifying rating variation comes from marginal differences in indicator pass rates. The bottom panel uses the position of the MSR binding indicator relative to the threshold as an instrument for the school rating. The resulting identifying rating variation comes from marginal differences in indicator size. Both panels control for the distance from the cutoff using a non-parametric local linear regression (LLR). Panel A includes estimates for bandwidths of 1,3 and 5 percentage points while panel B includes estimates for bandwidths of 1,3, and 5 students. What is clear from these estimates is that a negative shock at the *Acceptable-Unacceptable* boundary results in greater math gains the following year than shocks at other boundaries, and that a shock at the *Exemplary-Recognized* boundary may even lead to decreased mean test score gains. Due to a weak first stage effect at the Acceptable-Unacceptable boundary using the pass-rate discontinuity, the IV estimates blow-up and I do not

[18] They estimate that an 11 percentage point change in the proportion of students that are new to a school results in a 0.013 standard deviation decrease in achievement. My back of the envelope calculation assumes all spots vacated by non-reenrollees are filled by new students. In both the case where a campus is expanding and where a campus is contracting, the adverse mobility effect on achievement due to non-reenrollment would be less.

report them. Nevertheless, the reduced form effect at this boundary is two to three times larger than the effect at the *Recognized-Acceptable* boundary. The largest effects occur at the Acceptable-Unacceptable boundary using the MSR instrument where reduced form estimates of the effect of being treated with a low-performing evaluated indicator results in a 3-5 percentage point gain in the following years math test scores. While I prefer to emphasize the reduced form effects, when incorporating the probability of actually receiving the unacceptable rating, the IV estimates suggest a very large 10-20 percentage point gain in the pass-rate the year following a shock at the *Acceptable-Unacceptable* boundary. These effects can be seen in figure 7, which plot mean campus math pass-rate by the distance to the three rating discontinuities for both instruments. In essence, these figures are pictures of the reduced form effects. There is little discernable discontinuity. Table 6 and Figure 8 show that similar results are found for reading.

**7.2 School Fixed Effect**

The effects of ratings on the next year's campus-wide math pass-rate are given in table 7. Taking the lagged dependent variable OLS estimates and the gains estimates as bounds, exceptional schools show a 2.5 to 4.3 percentage point drop in the math pass-rate, while recognized drop between .7 to 2.5 percentage points. Unacceptable schools show an increase relative to acceptable schools of between 2.5 to 4.9 percentage points. Results for reading exhibit the same pattern but with smaller effects at each rating level.

These results are reflective of the adverse reenrollment response to a low rating or a negative rating shock and the fact that the official consequences associated with the rating system are concentrated at the unacceptable level. There is concern that the FE results may be partly driven by mean reversion (the degree to which requires further investigation). Nevertheless, taken together, the school fixed effects results and the regression discontinuity results provide two different types of evidence supporting the notion that schools do act on both the regulatory and the market based accountability incentives by improving math and reading achievement when assigned a low rating.

### 7.3 Heterogeneity by School Type: Charter vs. TPS [Future Work]

With greater operational flexibility, specifically in terms of staffing, and a more mobile student body, charter schools may exhibit a greater achievement response to a rating shock when compared to TPS.

## 8. Incoming student and sending campus profile results [Very Preliminary]

Until this point the focus has been on the response to ratings by schools on how they provide education and by families on whether they decide to switch schools. Selected characteristics of incoming students by school rating are plotted in figure 9. The profile of students new to a school can be viewed in terms of student level characteristics, and characteristics of the schools from which these students separated, what I will refer to as the sending, or prior, school. Presented are densities of the proportion low income, prior math test scores, prior school mean math achievement and prior school quality in terms of math value-added for all incoming students by receiving school rating. The mean incoming student profile broken down by various other characteristics illustrate the same monotonic pattern.[19] In terms of student characteristics, the proportion white and both math and reading prior achievement are monotonically increasing in the rating, while the proportion black, Hispanic, low income and LEP are monotonically decreasing in the rating of the receiving school. Both in terms of mean achievement and value-added, and in both mathematics and reading, higher rated schools receive students coming from monotonically better schools across. While lower rated schools tend to have larger incoming cohorts and draw disproportionately from charters, the pattern is not

[19] See table B1 in the appendix.

monotonic in the receiving school rating.  It is clear that there are significant differences between the profiles of new students across schools with different ratings.

**8.1 Regression Discontinuity.**

Having established the strong response to a student's own school rating, I now go on to the impact that ratings of other schools have on the way families exercise school choice. While these results are very preliminary, I will comment on certain notable outcomes. Table 8 presents the estimates of the effect of a rating shock on characteristics of the next year's incoming student profile for all schools. Unlike the student-level reenrollment regressions, few of the coefficients from these campus-level regressions are statistically significantly different from zero at any conventional level. The student demographic characteristic estimates in panel (A) of table show no particular difference in the profile of incoming students along any socio-economic dimension. Interestingly, evidence in panel (B) at the MSR discontinuity shows that while shocks at the lower two boundaries result in a lower achieving incoming student profile, a shock from exemplary results in a stronger incoming profile.

Table 9 provides estimates of the effects on the mean sending school quality. Shocks at the MSR discontinuity result in students being drawn from lower achieving campuses but interesting, these campuses have a higher value added. Shocks at the pass-rate discontinuity show less of an effect except at the *Exemplary-Recognized* boundary where the mean sending campus achievement shows a large and statistically significant improvement after a negative rating shock.

**8.2 School Fixed Effects**

Results of campus-level regressions of the incoming student profile are given in table 10. Panels (A) and (B) illustrate the effect of lagged ratings on the quality of schools from which a school attracts new students, the sending school quality. TPS specifications that include school fixed effects uniformly illustrate that the quality of the next year's incoming students' sending school, as measured by both value added (VA) to math scores and mean math achievement, increases monotonically with the receiving schools rating. Consistent with the notion that the TPS is always available as an outside option, this pattern remains but with smaller estimates across the board for the two-year ahead incoming class. For charters the picture is less clear. First, with charters only making up approximately 3% of the entire sample, all estimates lose statistical significance. Second, while the mean value added of the sending schools both one- and two-years ahead are unaffected by the lagged rating, the mean math scores at sending schools differ qualitatively if you look one-year ahead versus two. Nevertheless, whereas TPS exhibit a strong monotonic relationship between the sending school quality and ratings, families who opt to send their children to charters appear less influenced by the accountability rating that the school is assigned and more so by fixed unobserved characteristics that are specific to the school.

Where panels (A) and (B) explore the quality of sending schools, panels (C)-(E) look at differences in the demographic profile of the incoming students. The large differences between incoming cohorts in the odd numbered specifications all drop dramatically when school fixed effects are included. This suggests that unobserved characteristics like school geographic location and the quality of feeder schools are both correlated with ratings and school demographic profiles. While estimates are statistically insignificant for charter entrants, and small in magnitude for both charter and TPS entrants, some interesting patterns do emerge. Once school fixed effects are included higher ratings are still associated with having a lower proportion of incoming student that are less likely to be black, low income, or with limited English proficiency (LEP). In similar results that are not shown, a higher rating is associated with incoming students that are more likely to be white, and less likely to be of Hispanic origin.

In the most compelling results in this section, panels (F) and (G) compare the incoming students' prior year math and reading test scores. I estimate large differences by rating in mean prior math achievement for the on-year ahead profile for both charters and TPS (results are smaller but still large for reading). Between the exemplary and unacceptable ratings, TPS schools experience differences in incoming student quality of almost 0.1 standard deviations, while charters experience a 0.2 standard deviation difference. To put this effect in context, a .1 standard deviation in 4$^{th}$ grade achievement has been associated with a 10 student reduction in class size (Hanushek, Kain and Rivkin (2005)). Charter schools that receive an exemplary rating attract students who exhibit achievement profiles comparable to having come from schools where class sizes are over 10 students smaller than where unacceptable charters are able to draw students from. The notion that learning begets learning (Heckman (2000)) emphasizes the importance of this result. Within school, a higher rating is associated with higher achieving incoming students which will in turn promote further high achievement, which results in a higher ratings, and so on.

**9. RD Assumptions**

For the fuzzy RD estimates in the prior sections to take on a causal interpretation, the position of a campus relative to the discontinuities, the instruments, must satisfy the exclusion restrictions. This requires that the instruments must not affect outcomes other than through their effect on rating assignment. Theory provides no reason why a small difference in the marginal indicator's pass-rate or size would make a difference on future test score gains or family perceptions about schools other than through the rating mechanism. However, discontinuities in other observables across the boundaries may be indicative of differential sorting which would violate the exclusion restriction. Figures in appendix D illustrate the continuity of the mean campus characteristics and densities at the boundaries. In all figures, any heaping on either side of the boundary in panel (A) would be indicative of the possibility that schools may be able to select into or out of treatment. The remaining panels summarize the mean campus characteristics, demographics and quality across the treatment boundaries. These figures, and point estimates of discontinuities (not shown), provide no significant evidence that campuses differ discontinuously along any of these dimensions.[20]

Another common point of contention with RD designs is the bandwidth specification. In the pass-rate analyses I take the conservative approach of providing estimates for several bandwidths and I control non-parametrically using local liner regression within the bandwidth. LLR is preferred to kernel estimation, especially in when using the pass-rate instrument, when the running variable has a direct effect on the outcome independent of treatment. As would be expected with such a narrow bandwidth, unreported analyses using higher order polynomial controls for the running variable make little qualitative difference in the results. Due to the quantity of estimates I only report school choice estimates using a bandwidth of 3 students or percentage points. So far, I have not yet compared these estimates to the results using other bandwidths.

## 10. Conclusion

The school rating system in Texas is not unique in its reduction of a variety of continuous performance metrics into a single discrete quality measure. By unpacking the rating assignment system it is possible to exploit the cutoffs between rating levels to find a series of quasi-experiments that yield plausibly random rating assignments. The resulting causal estimates of the effects of rating, together with the broader fixed effect analyses, can provide policy guidance regarding the continuation, reduction or refinement of such systems.

The use of accountability ratings as a way of improving education is based on the notion that ratings can affect the behavior of the actors involved in education system. In this paper I

[20] Note that the discontinuities in the number of evaluated subgroups at the MSR boundaries is a direct consequence of treatment and not a potential violation of the RD assumptions. A discontinuity in the number of evaluated groups at the PR boundaries would have been a cause for concern.

supply evidence regarding how families and schools respond to ratings: Does a school's rating affect a family's decision on whether or not to switch schools? Do schools respond to low ratings by changing the way they provide education in ways that can result in higher achievement? Findings here are incomplete, so it is difficult to draw policy implications. Nevertheless, this study adds to the evidence indicating that accountability ratings can induce improvements for low performing schools while the role of ratings at higher achieving schools paints a less clear picture. Given the previous literatures findings regarding the adverse effects of student mobility on both the movers and their schoolmates, the fact that essentially meaningless shocks to information about quality can influence student mobility to such a degree suggests that a system that assigns ratings in a more continuous or meaningful manner may bring about achievement gains through the effect on student stability.

In closing, I would like to reemphasize the fact that the two different identifications strategies in this paper are based on very different sources of variation in ratings. The RD results, which are internally valid, but limited in terms of external validity, and the FE results which have more external validity but less claim to a causal interpretation, paint a consistent picture. The pure effect of being assigned a lower rating on a school's attractiveness to families is negative, while the effect on future achievement is positive.

**References**

Ball, Stephen J., Carol Vincent (1998). "'I Heard It on the Grapevine': 'Hot' Knowledge and School Choice." *British Journal of Sociology of Education*, Vol. 19, No. 3 (Sept), pp. 377-400.

Baude, Patrick L., Marcus Casey, Eric A. Hanushek and Steven G. Rivkin (2013) "Charter School Market Dynamics and School Quality." Working paper prepared for the CES-ifo conference on the economics of education, 2013.

Chiang, Hanley (2009). "How accountability pressure on failing schools affects student achievement." *Journal of Public Economics*, vol. 93, issue 9-10, pages 1045-1057.

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

Friesen, Jane, Mohsen Javdani, Justin Smith, and Simon Woodcock (2012). "How do school 'report cards' affect school choice decisions?" *Canadian Journal of Economics*, Vol. 45, No. 2 May: 784-807.

Hanushek, Eric A., John F. Kain and Steven G. Rivkin. (2004) "Disruption versus Tiebout improvement: the costs and benefits of switching schools," *Journal of Public Economics*, Vol. 88, pp. 1721–1746.

Hanushek, Eric A., John F. Kain and Steven G. Rivkin. (2005) "Teachers Schools and Academic Achievement," *Econometrica*, Vol. 73, No. 2, March, 417–458.

Hanushek, Eric A., John F. Kain, Steven G. Rivkin and Gregory F. Branch. (2007) "Charter School Quality and Parental Decision Making with School Choice." *Journal of Public Economics*, 91(5-6), June, pp. 823-848.

Hastings, J., and J. Weinstein (2008) "Information, school choice and academic achievement: evidence from two experiments", *Quarterly Journal of Economics* 123, 1373–414.

Heckman, J. J. (2000). *Invest in the very young*. Chicago: Ounce of Prevention Fund and the University of Chicago Harris School of Public Policy Studies.

Wooldridge, Jeffrey M., *Econometric Analysis of Cross Section and Panel Data,* MIT Press, 2002.

Imberman, Scott A. (2011), "Achievement and behavior in charter schools: Drawing a more complete picture." *The Review of Economics and Statistics*. May 2011, 93(2): 416–435.

Jacob, Brian, Lars Lefgren, and David Sims. (2010). "The Persistence of Teacher-Induced Learning Gains." *Journal of Human Resources* 45(4): 915-943.

Jin, Ginger Zhe, and Alan T. Sorenson (2006). "Information and consumer choice: The value of publicized health plan ratings." *Journal of Health Economics,* 25: 248–275.

Kain, John F. 2001. "The UTD Texas Schools Microdata Panel ( TSMP ): Its History, Use and Ways to Improve State Collection of Public School Data." Paper prepared for The Secretary's Forum on Research and Value-Added Assessment Data: U.S. Department of Education (December 5).

Kane, Thomas J., and Douglas O. Staiger (2002) "Racial Subgroup Rules in School Accountability Systems," Annual CREST conference, "Research Goes to School: Assessment, Accountability, and Improvement," UCLA.

Loeb, Susanna, Jon Valant, and Matt Kasman (2011). "Increasing choice in the market for schools: Recent reforms and their effects on student achievement", *National Tax Journal* March 2011, 64 (1), 141–164.

Meghir, Costas & Steven Rivkin (2011), "Econometric Methods for Research in Education", Chapter 1 In: Eric A. Hanushek, Stephen Machin and Ludger Woessmann, Editor(s), Handbook of the Economics of Education,, Volume 3, pages 1-87.

Mizala, Alejandra & Urquiola, Miguel (2013). "School markets: The impact of information approximating schools' effectiveness," *Journal of Development Economics*, Elsevier, vol. 103(C), pages 313-335.

Reardon, Sean F., and Joseph P. Robinson. 2012 "Regression discontinuity designs with multiple rating-score variables." *Journal of Research on Educational Effectiveness*. 5(1) 83-104.

Rockoff, Jonah, and Lesley J. Turner. (2010) "Short-run impacts of accountability on school quality," *American Economic Journal: Economic Policy*. 2, November: 119-147.

Rouse, Cecilia E., Jane Hannaway, Dan Goldhaber, and David Figlio (2013). "Feeling the Florida Heat? How Low-Performing Schools Respond to Voucher and Accountability Pressure." *American Economic Journal: Economic Policy*, 5(2): 251-81.

Sims, David P. (2013) "Can failure succeed? Using racial subgroup rules to analyze the effect of school accountability failure on student performance." *Economics of Education Review* 32, February, 262-274.

Todd, Petra E. and Kenneth I. Wolpin (2007). "The Production of Cognitive Achievement in Children: Home, School and Racial Test Score Gaps." *Journal of Human Capital* 1:1 (2007), 91–136.

van Dunk, Emily and Anneliese Dickman (2002). "School Choice Accountability: An Examination of Informed Consumers in Different Choice Programs." *Urban Affairs Review* Vol. 37, No. 6, July 2002: 844-856.

Panel A. No controls for campus quality Panel B. Math VA controls

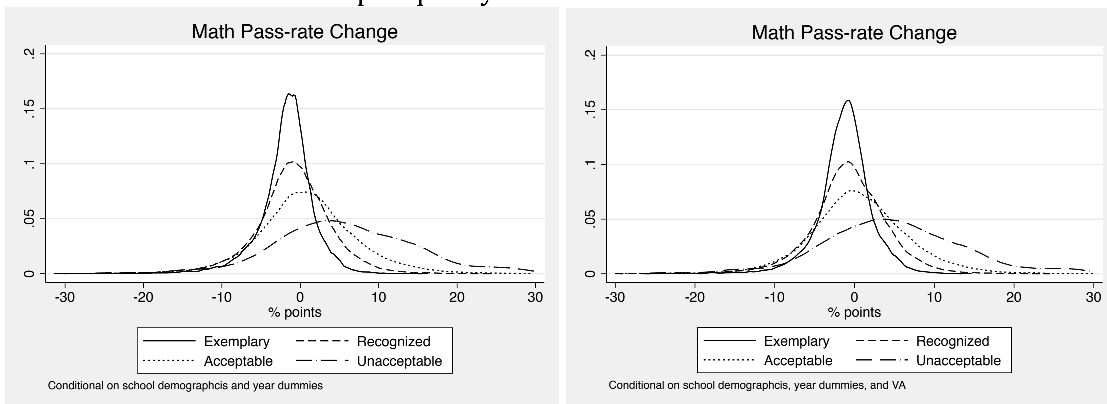



FIG 1. RESIDUAL PASS-RATE DISTRIBUTIONS BY LAGGED SCHOOL RATING.

*Notes:* Figures are based the residuals from regressions of one year ahead pass-rates gains on school demographic characteristics and year indicators. Panel B adds a linear control for campus quality using estimates of campus value added to math achievement.

Panel A. No controls for campus quality Panel B. Math VA controls

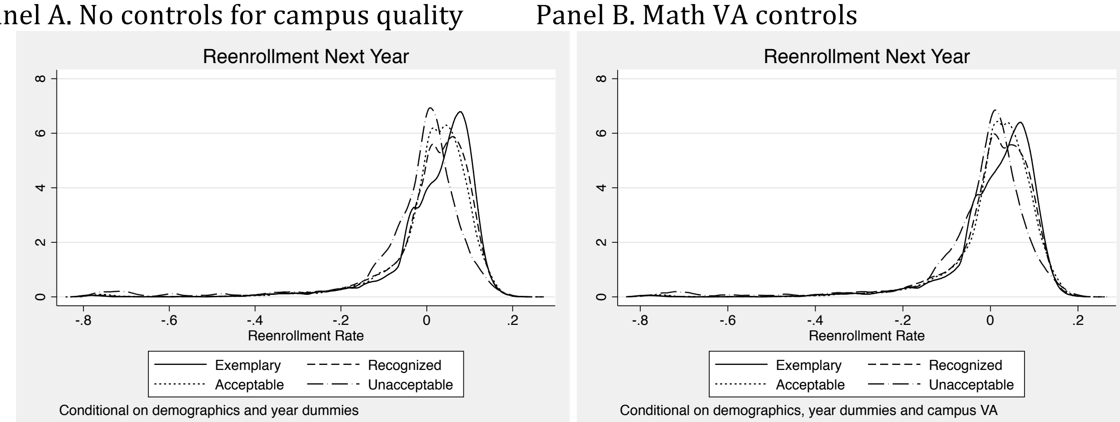



FIG 2. RESIDUAL REENROLLMENT DISTRIBUTIONS BY LAGGED SCHOOL RATING.

*Notes:* Figures are based the residuals from regressions of one year ahead reenrollment on controls for student demographics and year indicators. Panel B adds a linear control for campus quality using estimates of campus value added to math achievement.

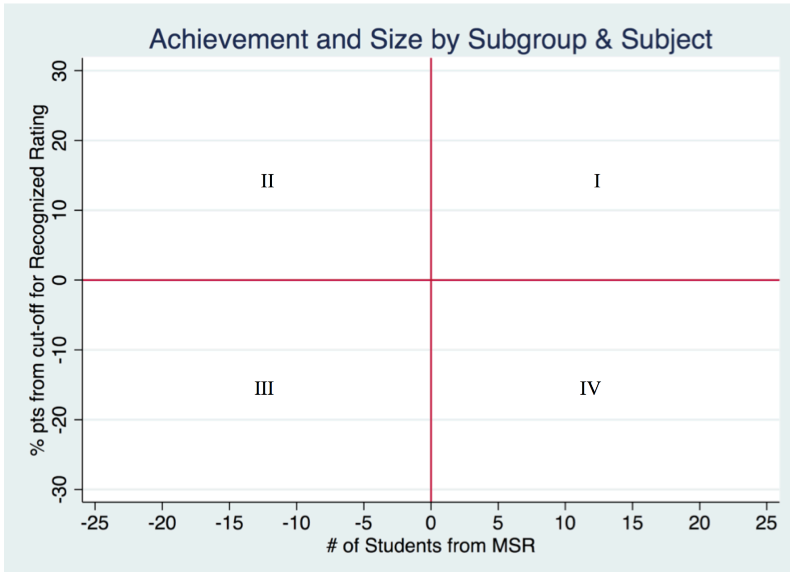
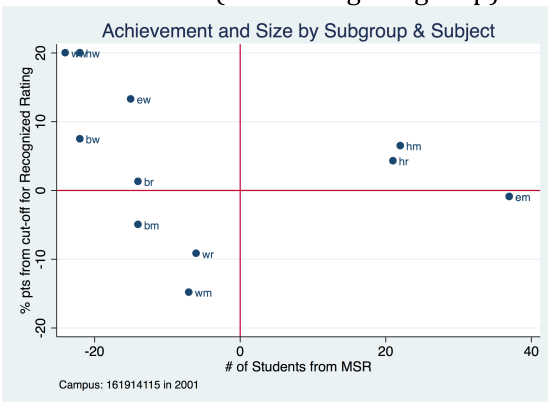


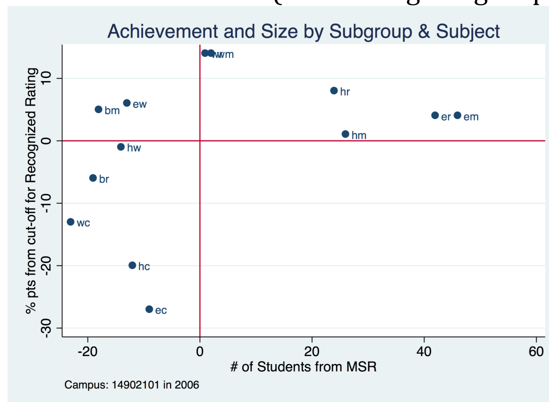

FIGURE 3. TREATMENT SPACE

*Notes*: This figure illustrates the evaluation status in the achievement-size space for each TAKS/TAAS indicator. The pass rate discontinuity compares schools across the I-IV frontier. The minimum size requirement discontinuity compares schools across the III-IV frontier.

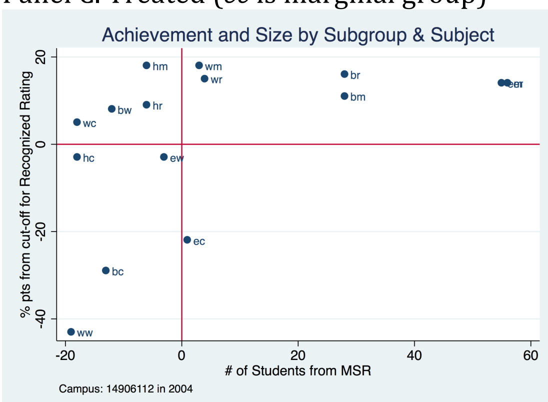
Panel A. Treated (*em* is marginal group)




Panel B. Untreated (*hm* is marginal group)




Panel C. Treated (*ec* is marginal group)




Panel D. Untreated (*hc* is marginal group)

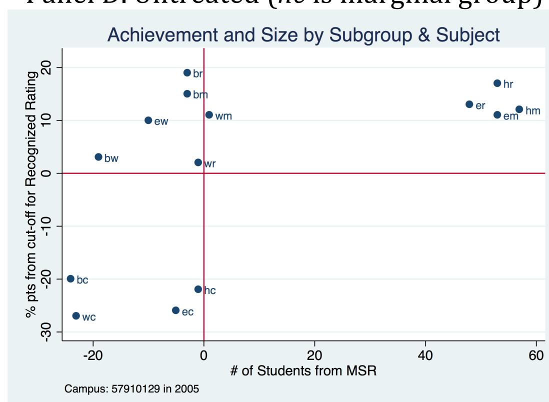
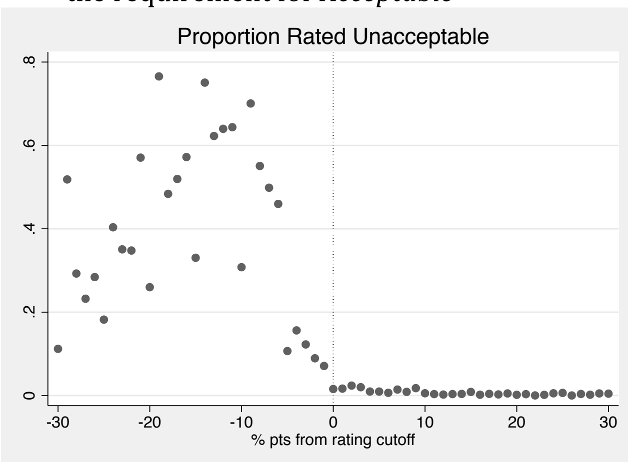


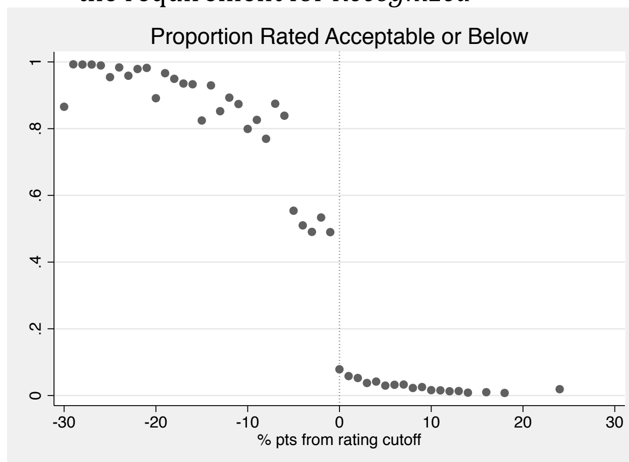

FIGURE 4. EXAMPLE CAMPUSES BY TREATMENT STATUS.

*Notes:* Examples of actual campuses that are treated and untreated. Panels (a) and (b) illustrate campuses within a small bandwidth of one percentage point of the pass-rate induced discontinuity. Panels (c) and (d) illustrate campuses within a bandwidth of one student of the minimum size requirement induced discontinuity.

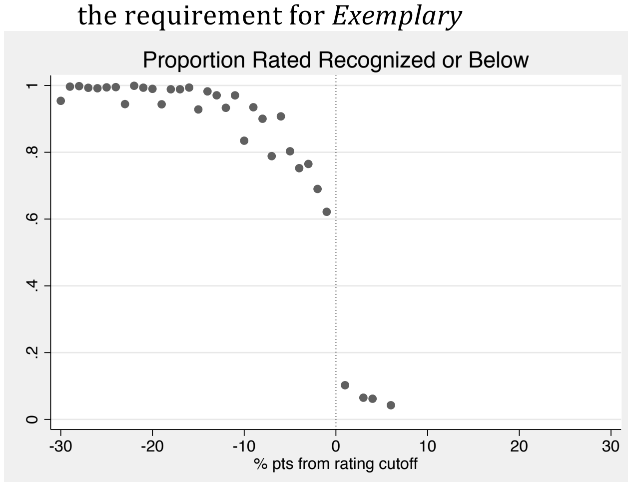
Panel A. Marginal group pass-rate distance from the requirement for *Acceptable*




Panel B. Marginal group pass-rate distance from the requirement for *Recognized*




Panel C. Marginal group pass-rate distance from the requirement for *Exemplary*



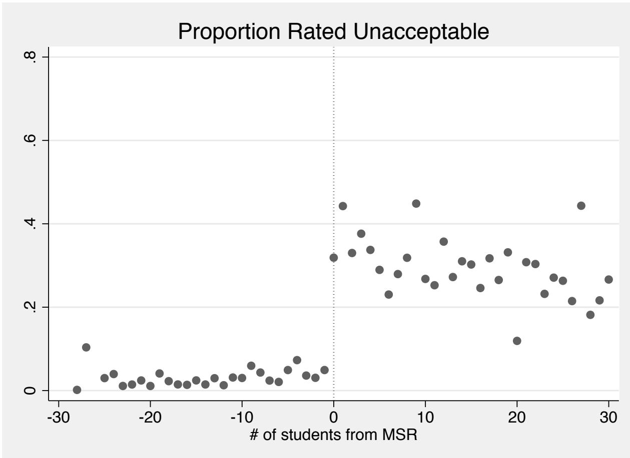

(c)

FIGURE 5. FIRST STAGES: GROUP PASS-RATE INSTRUMENTS

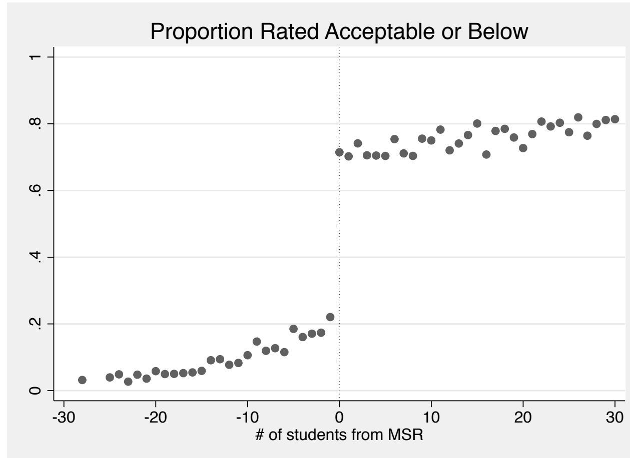
*Notes*: Each panel gives the proportion of schools rated below the specified rating level by the schools' marginal groups' percentage point distance from the required pass-rate for that rating. In this context, a school's marginal group is the lowest performing subgroup that is large enough to enter into the algorithm for determining the school's state accountability rating. Being to the *left* of the boundary corresponds to the presence of a low performing group that is subject to evaluation thus inducing treatment with a higher probability of receiving a lower rating.

Panel A. Marginal sub-*acceptable* group size

Panel B. Marginal sub-*recognized* group size distance from the MSR.




distance from the MSR.




Panel C. Marginal sub-*exemplary* group size
Distance from the MSR.

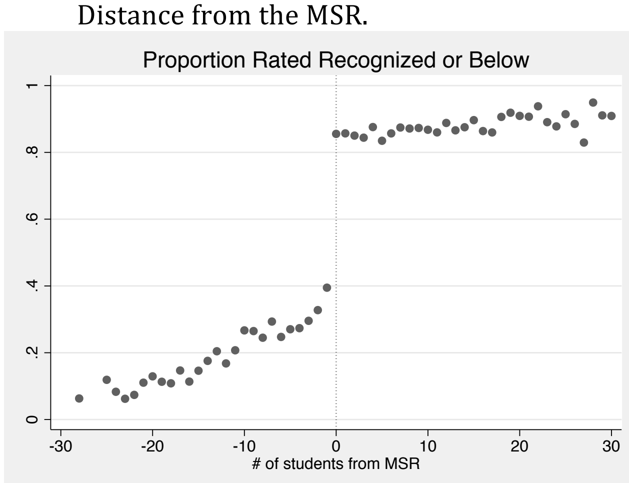
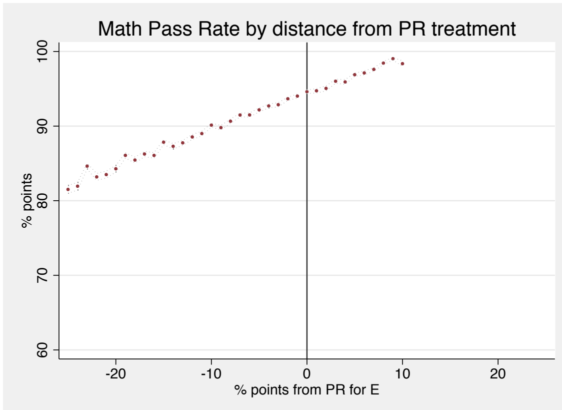
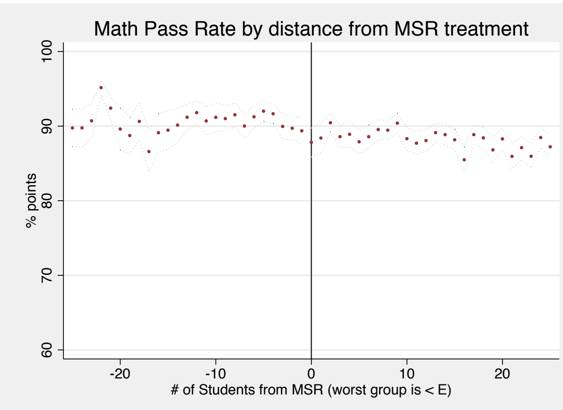



FIGURE 6. FIRST STAGES: GROUP SIZE INSTRUMENTS

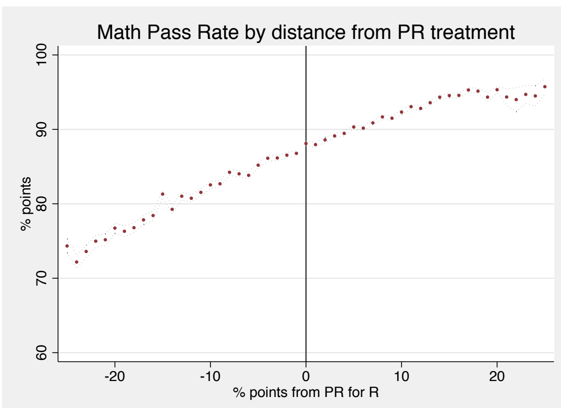
*Notes*: Each panel gives the proportion of schools rated below the specified rating level by the schools' marginal groups' number of student distance from the minimum size requirement for that group to be evaluated in that subject. In this context, a school's marginal group is the largest subgroup whose pass-rate is below the requirement for the specified rating. Being to the *right* of the boundary corresponds to the presence of a low performing group that is subject to evaluation thus inducing treatment with a higher probability of receiving a lower rating.



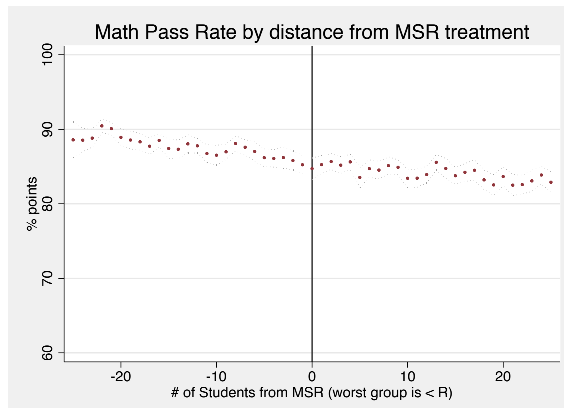

(a)



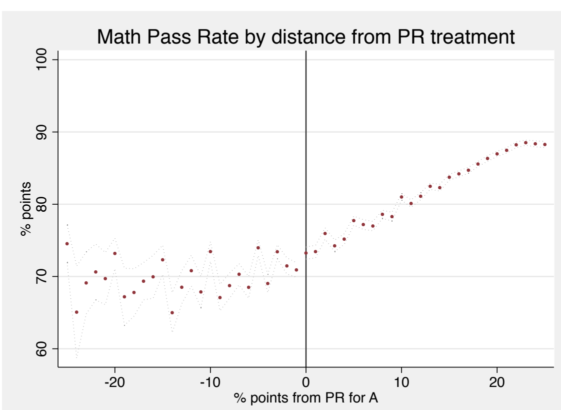

(b)



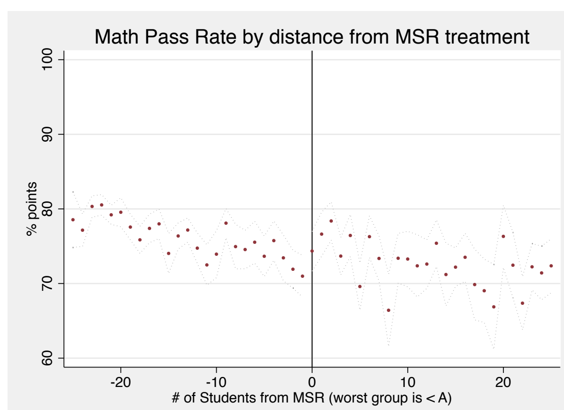

(c)




(d)



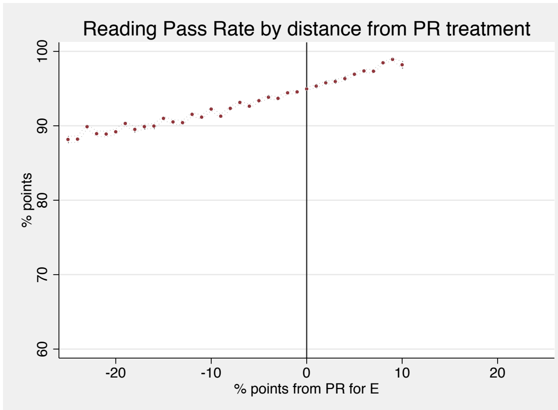

(e)



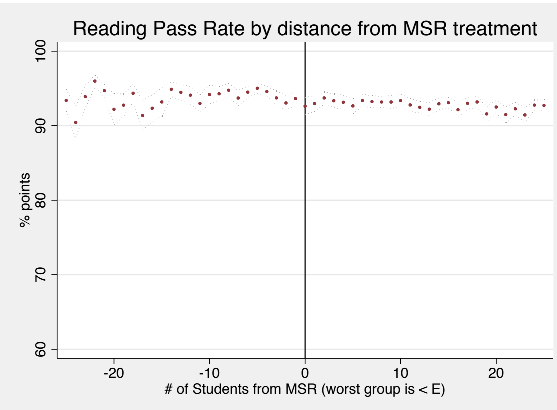

(f)

FIGURE 7. RD RESULTS – MATH PASS RATE

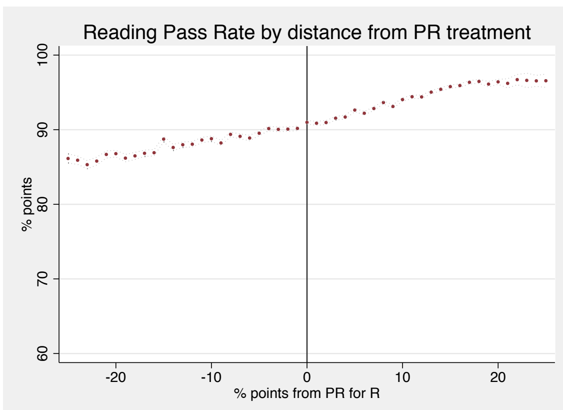
*Notes*: Mean MATH pass rate the following year by distance from discontinuity. Left hand figures correspond to pass-rate induced discontinuities, while right hand figures correspond to MSR induced discontinuities. Panels (a) and (b) are for the exceptional-recognized boundary, (c) and (d) are for the recognized-acceptable boundary, and (e) and (f) are for the acceptable-unacceptable boundary.



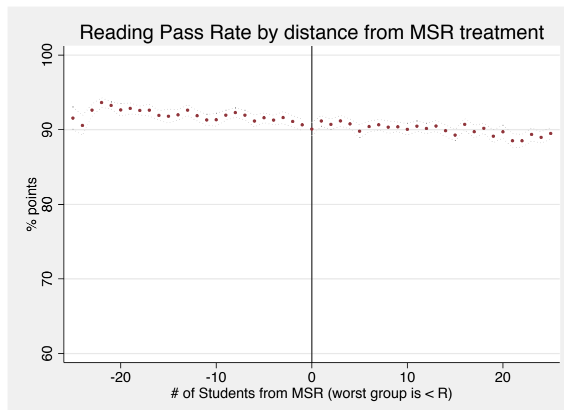

(a)



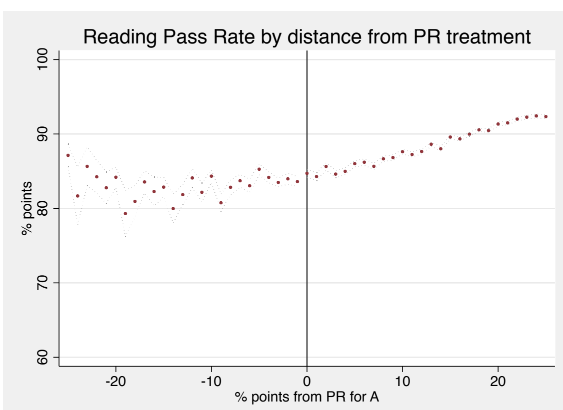

(b)



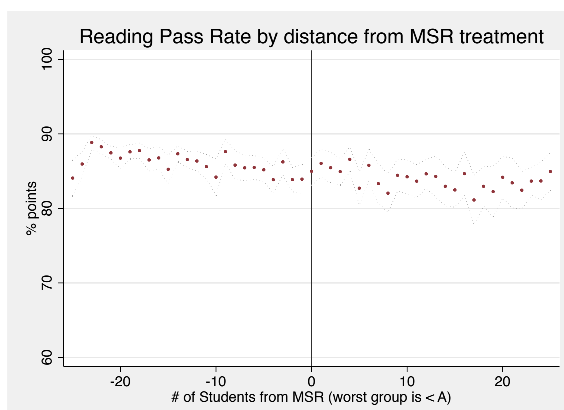

(c)




(d)




(e)

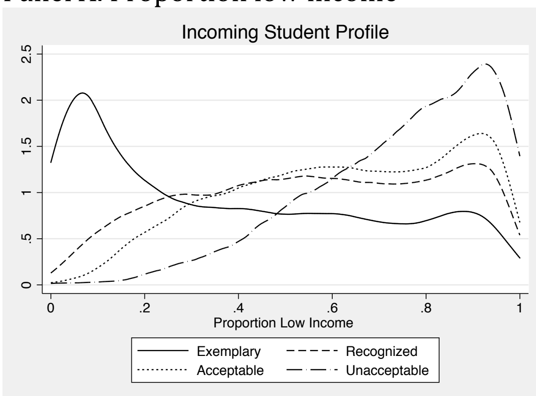



(f)

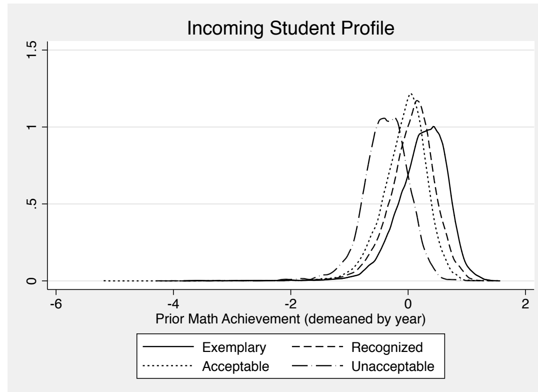
FIGURE 8. RD RESULTS – READING PASS RATE

*Notes*: Mean READING pass rate the following year by distance from discontinuity. Left hand figures correspond to pass-rate induced discontinuities, while right hand figures correspond to MSR induced discontinuities. Panels (a) and (b) are for the exceptional-recognized boundary, (c) and (d) are for the recognized-acceptable boundary, and (e) and (f) are for the acceptable-unacceptable boundary.

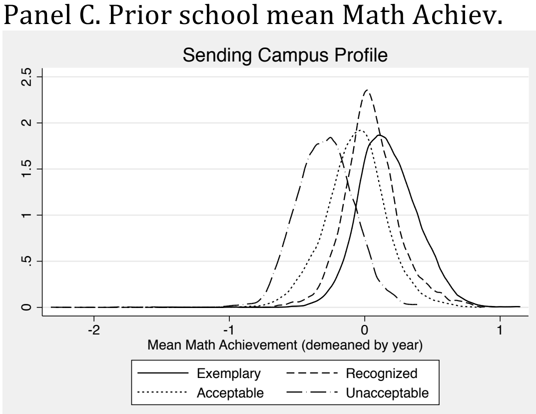
Panel A. Proportion low income




Panel B. Prior Math Achievement




Panel C. Prior school mean Math Achiev.




Panel D. Prior school Math VA

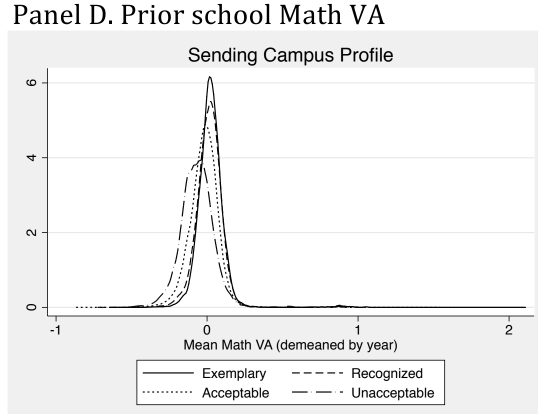



FIG 9. INCOMING STUDENT AND SENDING CAMPUS PROFILES BY PRIOR YEAR RATING.

*Notes:* Figures are based on raw incoming student characteristics. All panels with the exception of panel (A) are demeaned by year.

TABLE 1. TREATMENT MATRIX

| Quadrant | Pass Rate Shortfall? | Exceeds MSR? | Increased Probability of Rating Shock |
|---|---|---|---|
| I | N | Y | N |
| II | N | N | N |
| III | Y | N | N |
| IV | Y | Y | Y |

*Note*: Quadrants correspond to the treatment space depicted in figure 1.

TABLE 2. IV RESULTS – TPS REENROLLMENT

| | Pass Rate Discontinuity | | | MSR Discontinuity | | |
|---|---|---|---|---|---|---|
| | Exemplary to Recognized | Recognized to Acceptable | Acceptable to Unacceptable | Exemplary to Recognized | Recognized to Acceptable | Acceptable to Unacceptable |
| *Panel A. One Year Ahead* | | | | | | |
| Reduced Form | -0.003*** | -0.013*** | -0.016*** | -0.009** | -0.007** | 0.007 |
| | (0.001) | (0.001) | (0.001) | (0.004) | (0.003) | (0.005) |
| IV | -0.008*** | -0.049*** | | -0.024** | -0.018*** | 0.022 |
| | (0.003) | (0.003) | | (0.010) | (0.007) | -0.015 |
| Mean | 0.747 | 0.726 | 0.683 | 0.776 | 0.779 | 0.678 |
| Std Dev | 0.435 | 0.446 | 0.465 | 0.417 | 0.415 | 0.467 |
| N | 1,306,929 | 2,410,099 | 1,516,194 | 158,647 | 254,223 | 82,440 |
| *Panel B. Two Years Ahead* | | | | | | |
| Reduced Form | -0.008 | -0.030*** | -0.033*** | -0.021*** | -0.001 | -0.025 |
| | (0.002) | (0.001) | (0.002) | (0.006) | (0.005) | (0.007) |
| IV | -0.024*** | -0.120*** | | -0.053*** | -0.003 | -0.094*** |
| | (0.006) | (0.006) | | (0.016) | (0.013) | (0.027) |
| Mean | 0.522 | 0.502 | 0.472 | 0.629 | 0.595 | 0.505 |
| Std Dev | 0.5 | 0.5 | 0.499 | 0.483 | 0.491 | 0.5 |
| N | 617,566 | 1,205,907 | 902,930 | 79,813 | 135,158 | 44,932 |

*Notes:* These student level regressions control for student and campus characteristics, year indicators, and the campus level running variable using a local linear regression within a three percentage point or student bandwidth of the pass rate and MSR discontinuities respectively (qualitatively similar estimates result from narrower bandwidths). 'Unacceptable' IV results excluded for the pass-rate discontinuity due to weak first stages. Standard errors in parentheses are clustered at the campus level. *, **, and *** denote significance at the 10%, 5%, and 1% levels respectively.

TABLE 3. RD RESULTS – CHARTER REENROLLMENT

| | Pass Rate Discontinuity | | | MSR Discontinuity | | |
|---|---|---|---|---|---|---|
| | Exemplary to Recognized | Recognized to Acceptable | Acceptable to Unacceptable | Exemplary to Recognized | Recognized to Acceptable | Acceptable to Unacceptable |
| *Panel A. One Year Ahead* | | | | | | |
| Reduced Form | -0.040*** | -0.093*** | -0.037 | -0.087*** | -0.036 | -0.179*** |
| | (0.008) | (0.009) | (0.014) | (0.027) | (0.029) | (0.035) |
| IV | -0.166*** | -0.476*** | -0.179** | -1.461* | -0.054 | -0.811*** |
| | (0.033) | (0.047) | (0.070) | (0.772) | (0.043) | (0.190) |
| Mean | 0.529 | 0.568 | 0.513 | 0.589 | 0.654 | 0.551 |
| Std Dev | 0.499 | 0.495 | 0.5 | 0.492 | 0.476 | 0.497 |
| N | 34,652 | 34,502 | 19,285 | 5,395 | 5,838 | 5,333 |
| *Panel B. Two Years Ahead* | | | | | | |
| Reduced Form | -0.074*** | -0.138*** | -0.108 | -0.093*** | -0.141*** | -0.095** |
| | -0.009 | (0.009) | (0.017) | (0.032) | (0.034) | (0.037) |
| IV | | | | | | |
| Mean | 0.285 | 0.267 | 0.287 | 0.325 | 0.39 | 0.276 |
| Std Dev | 0.452 | 0.443 | 0.452 | 0.468 | 0.488 | 0.447 |
| N | 26,378 | 26,525 | 14,838 | 3,870 | 4,589 | 4,386 |

*Notes:* Each cell corresponds to a separate student-level regression. These regressions control for student and campus characteristics, year indicators, and the campus level running variable using a local linear regression within a three percentage point or student bandwidth of the pass rate and MSR discontinuities respectively (qualitatively similar estimates result from narrower bandwidths). Standard errors in parentheses are clustered at the campus level. *, **, and *** denote significance at the 10%, 5%, and 1% levels respectively.

TABLE 4. FE RESULTS – REENROLLMENT HETEROGENEITY: CHARTERS VS. TPS

| | 1yr Ahead Reenrollment | | | | 2yr Ahead Reenrollment | | | |
|---|---|---|---|---|---|---|---|---|
| | Charters | | TPS | | Charters | | TPS | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Exemplary | 0.070*** | 0.003 | 0.014*** | 0.010*** | 0.050*** | -0.019 | 0.017*** | 0.003* |
| | (0.015) | (0.015) | (0.002) | (0.001) | (0.015) | (0.015) | (0.002) | (0.002) |
| Recognized | 0.070*** | 0.013 | 0.003*** | 0.005*** | 0.070*** | 0.011 | 0.005*** | 0.003*** |
| | (0.014) | (0.012) | (0.001) | (0.001) | (0.016) | (0.014) | (0.002) | (0.001) |
| Unacceptable | -0.122*** | -0.010 | -0.037*** | -0.006** | -0.112*** | -0.029** | -0.049*** | -0.014*** |
| | (0.017) | (0.013) | (0.004) | (0.003) | (0.018) | (0.012) | (0.005) | (0.004) |
| Campus FE | No | Yes | No | Yes | No | Yes | No | Yes |
| Mean | 0.51 | | 0.652 | | 0.279 | | 0.43 | |
| Std Dev | 0.5 | | 0.476 | | 0.449 | | 0.495 | |
| N | 596,093 | | 42,657,760 | | 504,830 | | 29,640,911 | |

*Notes:* All regressions are at the student-level of 1 year- and 2 year ahead reenrollment indicators on school accountability rating dummies (excluded rating group is 'Acceptable'). All specifications include controls for student demographics and year indicators. Even numbered specifications include campus fixed effects. Standard errors in parentheses are clustered at the campus level. *, **, and *** denote significance at the 10%, 5%, and 1% levels respectively.

## TABLE 5. RD RESULTS AT EACH RATING BOUNDARY
## MATH PASS RATE

***Panel A - Pass Rate Boundary***

| | Exemplary-Recognized | | | Recognized-Acceptable | | | Acceptable-Unacceptable | | |
|---|---|---|---|---|---|---|---|---|---|
| *Bandwidth:* | 1 | 3 | 5 | 1 | 3 | 5 | 1 | 3 | 5 |
| OLS | -0.267 | -0.418*** | -0.326*** | -0.0491 | 0.537*** | 0.490*** | -0.951 | 1.108* | 1.018** |
| | (0.215) | (0.148) | (0.118) | (0.284) | (0.173) | (0.137) | (1.395) | (0.672) | (0.484) |
| Reduced Form | 0.0226 | 0.275 | 0.142 | 0.19 | 0.163 | 0.304 | 0.664 | 0.672 | 0.586 |
| | (0.178) | (0.227) | (0.166) | (0.193) | (0.251) | (0.185) | (0.426) | (0.538) | (0.377) |
| IV | 0.0627 | 0.875 | 0.407 | 0.78 | 0.686 | 1.26 | | | |
| | (0.492) | (0.726) | (0.477) | (0.791) | (1.052) | (0.769) | | | |
| Mean | 94.28 | 94.04 | 93.84 | 87.55 | 87.39 | 87.44 | 72.37 | 73.3 | 73.49 |
| Std Dev | 4.408 | 4.681 | 4.819 | 6.949 | 7.115 | 7.249 | 12.47 | 12.37 | 12.22 |
| First Stage F | 344.5 | 193.2 | 466 | 348.9 | 189.3 | 373 | 2.513 | 0.16 | 0.00456 |
| N | 1,495 | 4,404 | 7,136 | 2,379 | 6,822 | 11,366 | 1,097 | 3,114 | 5,164 |

***Panel B - MSR Boundary***

| | Exemplary-Recognized | | | Recognized-Acceptable | | | Acceptable-Unacceptable | | |
|---|---|---|---|---|---|---|---|---|---|
| *Bandwidth:* | 1 | 3 | 5 | 1 | 3 | 5 | 1 | 3 | 5 |
| OLS | 1.034 | 0.405 | 0.0825 | 0.436 | 0.346 | 0.00494 | 0.74 | 0.916 | 0.93 |
| | (1.146) | (0.560) | (0.409) | (0.721) | (0.443) | (0.329) | (2.069) | (1.198) | (0.903) |
| Reduced Form | 0.0682 | -0.149 | 0.159 | 0.666 | 1.072 | 0.55 | 3.174** | 4.834*** | 2.806** |
| | (0.905) | (0.979) | (0.667) | (0.645) | (0.765) | (0.551) | (1.499) | (1.820) | (1.257) |
| IV | 0.252 | -0.842 | 0.646 | 1.824 | 3.537 | 1.488 | 15.63* | 21.97* | 11.45** |
| | (3.263) | (5.482) | (2.692) | (1.749) | (2.562) | (1.494) | (8.838) | (11.38) | (5.655) |
| Mean | 88.46 | 89.32 | 89.74 | 85 | 85.5 | 85.61 | 72.75 | 74.05 | 74.31 |
| Std Dev | 13.35 | 12.42 | 11.78 | 11.2 | 10.65 | 10.24 | 11.51 | 12.61 | 12.43 |
| First Stage F | 34.36 | 8.98 | 33.50 | 80.31 | 39.45 | 107.80 | 7.47 | 6.91 | 18.03 |
| N | 271 | 878 | 1,445 | 416 | 1,259 | 2,126 | 105 | 344 | 583 |

*Notes*: Effect of a negative rating shock on MATH performance the next year. Panel A instruments for the shock using treatment across the pass rate threshold. Panel B instruments for the shock using treatment across the MSR threshold. All specifications control for the lagged pass rate, campus demographics and year fixed effects. The distance from treatment enters non-parametrically using local linear regression (LLR) within bandwidths of 1, 3 or 5 percentage points or students. LLR trend coefficients are allowed to differ across the threshold. Panel A 'acceptable-unacceptable' IV coefficients omitted due to weak first stages. Clustered standard errors at the campus level in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels respectively.

TABLE 6. RESULTS – MATH PASS-RATE

| | Lagged Dep Variable - OLS | | Lagged Dep Variable - IV | | Gains | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| *Lagged Rating* | | | | | | |
| Exceptional | -0.824*** | -2.560*** | -1.091*** | -2.338*** | -2.902*** | -4.279*** |
| | (0.078) | (0.099) | (0.094) | (0.122) | (0.064) | (0.118) |
| Recognized | -0.276*** | -0.734*** | -0.627*** | -0.889*** | -1.945*** | -2.456*** |
| | (0.059) | (0.067) | (0.071) | (0.083) | (0.053) | (0.081) |
| Unacceptable | 2.195*** | 2.540*** | 2.218*** | 2.241*** | 4.293*** | 4.857*** |
| | (0.230) | (0.238) | (0.240) | (0.256) | (0.230) | (0.295) |
| Campus FE | N | Y | N | Y | N | Y |
| Mean | 82.64 | 82.64 | 83.58 | 83.58 | 1.892 | 1.892 |
| SD | 12.86 | 12.86 | 12.06 | 12.06 | 5.081 | 5.081 |
| First Stage F | . | . | 3609.8 | 1100 | . | . |
| N | 59,561 | 59,561 | 51,303 | 51,010 | 59,561 | 59,561 |

*Notes:* Estimates from regressions on lagged rating dummies with *acceptable* the excluded group. All regressions are enrollment weighted and include year dummies and campus demographics. IV columns instrument for lagged pass rates with twice lagged pass rates. Even numbered columns control for campus fixed effects. Clustered standard errors at the campus level in parentheses. *, **, and *** denote significance at the 5%, 1%, and 0.1% levels respectively.

TABLE 7. RD RESULTS AT EACH RATING BOUNDARY
READING PASS RATE

***Panel A - Pass Rate Boundary***

| | Exemplary-Recognized | | | Recognized-Acceptable | | | Acceptable-Unacceptable | | |
|---|---|---|---|---|---|---|---|---|---|
| *Bandwidth:* | 1 | 3 | 5 | 1 | 3 | 5 | 1 | 3 | 5 |
| OLS | -0.525*** | -0.281** | -0.274*** | -0.254 | 0.0527 | 0.0932 | 0.473 | 0.465 | 0.311 |
| | (0.185) | (0.120) | (0.097) | (0.213) | (0.134) | (0.104) | (0.964) | (0.475) | (0.353) |
| Reduced Form | -0.246 | -0.102 | -0.249* | 0.0289 | 0.108 | 0.169 | 0.364 | 0.568 | 0.253 |
| | (0.155) | (0.185) | (0.137) | (0.145) | (0.194) | (0.141) | (0.295) | (0.381) | (0.275) |
| IV | -0.651 | -0.312 | -0.698* | 0.116 | 0.442 | 0.688 | | | |
| | (0.408) | (0.563) | (0.383) | (0.580) | (0.796) | (0.575) | | | |
| Mean | 94.71 | 94.64 | 94.51 | 90.62 | 90.52 | 90.61 | 84.28 | 84.41 | 84.57 |
| Std Dev | 3.771 | 3.765 | 3.932 | 5.311 | 5.519 | 5.551 | 7.877 | 7.74 | 7.683 |
| First Stage F | 356 | 200.1 | 463.4 | 362.9 | 196 | 381.5 | 2.852 | 0.334 | 0.00452 |
| N | 1404 | 4211 | 6832 | 2357 | 6774 | 11285 | 1095 | 3111 | 5160 |

***Panel B - MSR Boundary***

| | Exemplary-Recognized | | | Recognized-Acceptable | | | Acceptable-Unacceptable | | |
|---|---|---|---|---|---|---|---|---|---|
| *Bandwidth:* | 1 | 3 | 5 | 1 | 3 | 5 | 1 | 3 | 5 |
| OLS | 0.276 | -0.243 | -0.454 | -0.0255 | -0.0582 | -0.328 | 0.905 | 0.744 | 0.265 |
| | (0.777) | (0.414) | (0.316) | (0.501) | (0.314) | (0.232) | (1.322) | (0.816) | (0.648) |
| Reduced Form | 0.347 | 0.17 | 0.48 | 0.668 | 1.038* | 0.557 | 1.750* | 2.913** | 1.24 |
| | (0.611) | (0.725) | (0.514) | (0.455) | (0.548) | (0.394) | (0.952) | (1.226) | (0.905) |
| IV | 1.238 | 0.924 | 1.938 | 1.826 | 3.492* | 1.537 | 10.87 | 16.92 | 5.771 |
| | (2.133) | (3.929) | (2.104) | (1.247) | (1.927) | (1.098) | (7.142) | (10.40) | (4.421) |
| Mean | 93.06 | 93.26 | 93.57 | 90.39 | 90.9 | 91.02 | 84.48 | 85.04 | 85.02 |
| Std Dev | 7.6 | 7.299 | 6.785 | 6.865 | 6.419 | 6.155 | 7.959 | 8.178 | 8.011 |
| First Stage F | 36.99 | 9.207 | 32.25 | 76.96 | 36.77 | 100.5 | 4.694 | 4.303 | 13.71 |
| N | 267 | 844 | 1369 | 413 | 1248 | 2110 | 105 | 343 | 582 |

*Notes*: Effect of a negative rating shock on READING performance the next year. Panel A instruments for the shock using treatment across the pass rate threshold. Panel B instruments for the shock using treatment across the MSR threshold. All specifications control for the lagged pass rate, campus demographics and year fixed effects. The distance from treatment enters non-parametrically using local linear regression (LLR) within bandwidths of 1, 3 or 5 percentage points or students. LLR trend coefficients are allowed to differ across the threshold. Panel A 'acceptable-unacceptable' IV coefficients omitted due to weak first stages. Clustered standard errors in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels respectively.

TABLE 8. RD RESULTS AT EACH RATING BOUNDARY
INCOMING COHORT PROFILE

| | Pass Rate Discontinuity | | | MSR Discontinuity | | |
|---|---|---|---|---|---|---|
| | Exemplary to Recognized | Recognized to Acceptable | Acceptable to Unacceptable | Exemplary to Recognized | Recognized to Acceptable | Acceptable to Unacceptable |
| *Panel A. Demographics* | | | | | | |
| White | 0.010 | -0.002 | | 0.005 | 0.018 | 0.072 |
| | 0.006 | (0.005) | | (0.020) | (0.013) | (0.048) |
| Mean/SD | .449/.302 | .365/.283 | | .603/.224 | .513/.262 | .334/.273 |
| N/F | 8616/701.4 | 11457/619 | | 1767/54.4 | 2154/112 | 508/12.53 |
| Black | -0.008* | 0.002 | | -0.009 | -0.027** | 0.003 |
| | (0.005) | (0.005) | | (0.012) | (0.011) | (0.266) |
| Mean/SD | .107/.147 | .140/.171 | | .106/.125 | .142/.153 | .224/.241 |
| N/F | 8616/701.4 | 11457/619 | | 1767/54.4 | 2154/112 | 508/12.53 |
| Hispanic | -0.003 | -0.007 | | 0.013 | 0.004 | -0.052 |
| | (0.006) | (0.005) | | (0.017) | (0.013) | (0.047) |
| Mean/SD | .396/.317 | .457/.308 | | .234/.194 | .307/.228 | .426/.270 |
| N/F | 8616/701.4 | 11457/619 | | 1761/54.4 | 2154/112 | 508/12.53 |
| Limited Eng | 0.002 | -0.005 | | -0.004 | 0.016 | -0.010 |
| | (0.008) | (0.009) | | (0.015) | (0.012) | (0.069) |
| Mean/SD | .164/.192 | .187/.201 | | .089/.116 | .122/.146 | .164/.173 |
| N/F | 8616/701.4 | 11456/619 | | 1767/54.37 | 2154/112 | 508/12.5 |
| Low Income | -0.000 | 0.000 | | 0.021 | 0.035* | 0.024 |
| | (0.007) | (0.007) | | (0.025) | (0.018) | (0.061) |
| Mean/SD | .485/.290 | .567/.263 | | .344/.236 | .491/.244 | .679/.207 |
| N/F | 8616/701.4 | 11457/619 | | 1767/54.4 | 2154/112 | 508/12.5 |
| *Panel B. Prior Achievement* | | | | | | |
| Math | 0.018 | 0.000 | | 0.159 | -0.129* | -0.118 |
| | (0.035) | (0.034) | | (0.100) | (0.075) | (0.234) |
| Mean/SD | .056/.412 | -.053/.421 | | .112/.433 | -.049/.417 | -.308/.436 |
| N/F | 8101/642 | 10844/569 | | 1652/55.1 | 2043/110 | 484/12.6 |
| Reading | 0.038 | 0.032 | | 0.037 | -0.148* | -0.125 |
| | (0.034) | 0.035 | | (0.093) | (0.077) | (0.226) |
| Mean/SD | .044/.418 | -.061/0.416 | | .119/.438 | -.030/.435 | -0.69248826 |
| N/F | 8106/641 | 10851/570 | | 1653/54.9 | 2045/109 | 484/12.6 |

*Notes:* Each cell corresponds to a separate campus-level regression. All regressions control for campus demographics, year indicators, and the running variable using a local linear regression within a three percentage point or student bandwidth of the pass rate and MSR discontinuities respectively. Pass-rate discontinuity estimates at the acceptable-unacceptable boundary omitted due to weak first stages. Statistics include the mean and standard deviation of the dependent variable, sample size, and the first-stage F-statistic for weak instruments. Standard errors in parentheses are clustered at the campus level. . *, **, and *** denote significance at the 10%, 5%, and 1% levels respectively.

TABLE 9. RD RESULTS AT EACH RATING BOUNDARY
SENDING CAMPUS PROFILE

| | Pass Rate Discontinuity | | | MSR Discontinuity | | |
|---|---|---|---|---|---|---|
| | Exemplary to Recognized | Recognized to Acceptable | Acceptable to Unacceptable | Exemplary to Recognized | Recognized to Acceptable | Acceptable to Unacceptable |
| *Panel A. Mean Campus Achievement* | | | | | | |
| Math | 0.050*** | 0.005 | | -0.081 | -0.115* | 0.071 |
| | (0.019) | (0.018) | | (0.128) | (0.060) | (0.116) |
| Mean/SD | .099/.220 | .023/.220 | | .120/.302 | .040/.268 | -.146/.219 |
| N/F | 8613/701 | 11454/619 | | 1767/54.4 | 2154/113 | 508/12.5 |
| Reading | 0.045*** | -0.002 | | -0.064 | -0.086** | 0.071 |
| | (0.016) | (0.015) | | (0.086) | (0.042) | (0.100) |
| Mean/SD | .083/.216 | .009/.210 | | .136/.236 | .050/.217 | -.134/.209 |
| N/F | 8613/701 | 11454/619 | | 1767/54.4 | 2154/113 | 508/12.5 |
| *Panel B. Campus Value Added* | | | | | | |
| Math | 0.008 | -0.022 | | 0.032 | 0.016 | 0.100 |
| | (0.016) | (0.014) | | (0.066) | (0.020) | (0.083) |
| Mean/SD | .054/.109 | .039/.111 | | .058/.155 | .024/.105 | -.017/.123 |
| N/F | 7306/552 | 9946/410 | | 1403/41.7 | 1790/86.9 | 417/12.5 |
| Reading | 0.009 | -0.030** | | 0.033 | 0.008 | .0.100 |
| | (0.015) | (0.014) | | (0.066) | (0.018) | (0.070) |
| Mean/SD | .031/.104 | .017/.108 | | .043/.163 | .003/.103 | -.028/.112 |
| N/F | 7306/552 | 9946/410 | | 1403/41.7 | 1790/86.9 | 417/12.53 |

*Notes:* Each cell corresponds to a separate campus-level regression. All regressions control for campus demographics, year indicators, and the running variable using a local linear regression within a three percentage point or student bandwidth of the pass rate and MSR discontinuities respectively. Pass-rate discontinuity estimates at the acceptable-unacceptable boundary omitted due to weak first stages. Statistics include the mean and standard deviation of the dependent variable, sample size, and the first-stage F-statistic for weak instruments. Standard errors in parentheses are clustered at the campus level. . *, **, and *** denote significance at the 10%, 5%, and 1% levels respectively.

TABLE 10. FE RESULTS – INCOMING STUDENT CHARACTERISTICS: CHARTERS VS. TPS

| | One Year Ahead | | | | Two Years Ahead | | | |
|---|---|---|---|---|---|---|---|---|
| | Charter | | TPS | | Charter | | TPS | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| *A. Sending School Quality: Mean Math Achievement* | | | | | | | | |
| Exemplary | 0.163*** | -0.016 | 0.254*** | 0.052*** | 0.181*** | 0.021 | 0.250*** | 0.036*** |
| | (0.031) | (0.024) | (0.004) | (0.002) | (0.035) | (0.024) | (0.005) | (0.002) |
| Recognized | 0.070*** | -0.010 | 0.133*** | 0.032*** | 0.095*** | 0.014 | 0.128*** | 0.023*** |
| | (0.020) | (0.013) | (0.003) | (0.001) | (0.022) | (0.016) | (0.003) | (0.002) |
| Unacceptable | -0.108*** | -0.020 | -0.195*** | -0.029*** | -0.087*** | -0.010 | -0.178*** | -0.003 |
| | (0.024) | (0.017) | (0.007) | (0.005) | (0.023) | (0.015) | (0.007) | (0.005) |
| Campus FE | No | Yes | No | Yes | No | Yes | No | Yes |
| Mean | -0.161 | | -0.011 | | -0.160 | | -0.012 | |
| Std Dev | 0.248 | | 0.224 | | 0.241 | | 0.224 | |
| N | 1,457 | | 80,031 | | 1,180 | | 72,176 | |
| *B. Sending School Quality: Math Value Added* | | | | | | | | |
| Exemplary | 0.016 | 0.006 | 0.044*** | 0.032*** | 0.023* | 0.004 | 0.037*** | 0.017*** |
| | (0.012) | (0.019) | (0.002) | (0.002) | (0.013) | (0.020) | (0.002) | (0.002) |
| Recognized | 0.012 | 0.001 | 0.030*** | 0.021*** | 0.021** | 0.006 | 0.024*** | 0.010*** |
| | (0.008) | (0.010) | (0.001) | (0.001) | (0.009) | (0.013) | (0.001) | (0.001) |
| Unacceptable | -0.00587 | -0.022 | -0.027*** | -0.022*** | 0.004 | -0.002 | -0.016*** | -0.002 |
| | (0.013) | (0.015) | (0.004) | (0.004) | (0.011) | (0.015) | (0.004) | (0.004) |
| Campus FE | No | Yes | No | Yes | No | Yes | No | Yes |
| Mean | 0.024 | | 0.017 | | 0.025 | | 0.019 | |
| Std Dev | 0.155 | | 0.108 | | 0.152 | | 0.11 | |
| N | 1,435 | | 67,810 | | 1,171 | | 66,148 | |

*Continued*

| | One Year Ahead | | | | Two Years Ahead | | | |
|---|---|---|---|---|---|---|---|---|
| | Charter | | TPS | | Charter | | TPS | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| *C. Demographics: Proportion Low Income* | | | | | | | | |
| Exemplary | -0.159*** | -0.019 | -0.277*** | -0.009*** | -0.167*** | -0.045 | -0.284*** | -0.011*** |
| | (0.039) | (0.022) | (0.005) | (0.002) | (0.046) | (0.030) | (0.006) | (0.002) |
| Recognized | -0.100*** | -0.012 | -0.114*** | -0.004*** | -0.122*** | -0.028 | -0.117*** | -0.005*** |
| | (0.025) | (0.017) | (0.003) | (0.001) | (0.027) | (0.021) | (0.003) | (0.001) |
| Unacceptable | 0.094*** | -0.036* | 0.106*** | 0.000 | 0.106*** | -0.025 | 0.113*** | 0.004 |
| | (0.025) | (0.022) | (0.006) | (0.003) | (0.028) | (0.024) | (0.006) | (0.003) |
| Campus FE | No | Yes | No | Yes | No | Yes | No | Yes |
| Mean | 0.64 | | 0.585 | | 0.649 | | 0.589 | |
| Std Dev | 0.311 | | 0.261 | | 0.311 | | 0.259 | |
| N | 1460 | | 80073 | | 1182 | | 72217 | |
| *D. Demographics: Proportion Black* | | | | | | | | |
| Exemplary | -0.151*** | -0.004 | -0.094*** | -0.006*** | -0.153*** | -0.006 | -0.092*** | -0.007*** |
| | (0.041) | (0.012) | (0.004) | (0.001) | (0.046) | (0.013) | (0.004) | (0.001) |
| Recognized | -0.079** | -0.005 | -0.060*** | -0.003*** | -0.080** | -0.010 | -0.058*** | -0.003*** |
| | (0.031) | (0.008) | (0.003) | (0.001) | (0.034) | (0.009) | (0.003) | (0.001) |
| Unacceptable | 0.051* | 0.002 | 0.150*** | -0.004** | 0.042 | 0.007 | 0.141*** | -0.005*** |
| | (0.031) | (0.010) | (0.009) | (0.002) | (0.032) | (0.008) | (0.009) | (0.002) |
| Campus FE | No | Yes | No | Yes | No | Yes | No | Yes |
| Mean | 0.337 | | 0.143 | | 0.341 | | 0.142 | |
| Std Dev | 0.338 | | 0.189 | | 0.336 | | 0.188 | |
| N | 1460 | | 80073 | | 1182 | | 72217 | |
| *E. Demographics: Proportion LEP* | | | | | | | | |
| Exemplary | -0.015 | -0.0202 | -0.0811*** | -0.00223* | -0.016 | -0.0325 | -0.0853*** | -0.00220* |
| | (0.025) | (0.017) | (0.003) | (0.001) | (0.029) | (0.022) | (0.004) | (0.001) |
| Recognized | -0.00613 | -0.0123 | -0.0344*** | 0.00203** | -0.00696 | -0.0105 | -0.0395*** | -0.00178** |
| | (0.017) | (0.011) | (0.002) | (0.001) | (0.021) | (0.014) | (0.002) | (0.001) |
| Unacceptable | -0.0153 | -0.0138 | 0.0212*** | -0.000948 | -0.00946 | -0.00395 | 0.0240*** | 0.000419 |
| | (0.014) | (0.012) | (0.005) | (0.002) | (0.014) | (0.010) | (0.005) | (0.003) |
| Campus FE | No | Yes | No | Yes | No | Yes | No | Yes |
| Mean | 0.125 | | 0.165 | | 0.13 | | 0.168 | |
| Std Dev | 0.208 | | 0.195 | | 0.212 | | 0.196 | |
| N | 1460 | | 80073 | | 1182 | | 72217 | |

*Continued*

| | One Year Ahead | | | | Two Years Ahead | | | |
|---|---|---|---|---|---|---|---|---|
| | Charter | | TPS | | Charter | | TPS | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| *F: Student Achievement: Math* | | | | | | | | |
| Exemplary | 0.508*** | 0.119** | 0.414*** | 0.066*** | 0.424*** | -0.019 | 0.402*** | 0.039*** |
| | (0.050) | (0.060) | (0.007) | (0.006) | (0.069) | (0.080) | (0.008) | (0.006) |
| Recognized | 0.286*** | 0.079 | 0.203*** | 0.043*** | 0.266*** | 0.017 | 0.195*** | 0.028*** |
| | (0.041) | (0.049) | (0.004) | (0.004) | (0.049) | (0.046) | (0.005) | (0.004) |
| Unacceptable | -0.267*** | -0.079 | -0.329*** | -0.053*** | -0.169*** | 0.033 | -0.301*** | -0.015 |
| | (0.044) | (0.050) | (0.013) | (0.010) | (0.048) | (0.055) | (0.013) | (0.011) |
| Campus FE | No | Yes | No | Yes | No | Yes | No | Yes |
| Mean | -0.325 | | -0.100 | | -0.341 | | -0.104 | |
| Std Dev | 0.591 | | 0.452 | | 0.582 | | 0.452 | |
| N | 1409 | | 75850 | | 1143 | | 68454 | |
| *G. Student Achievement: Reading* | | | | | | | | |
| Exemplary | 0.428*** | 0.057 | 0.383*** | 0.030*** | 0.341*** | -0.053 | 0.385*** | 0.025*** |
| | (0.048) | (0.059) | (0.007) | (0.006) | (0.070) | (0.103) | (0.008) | (0.006) |
| Recognized | 0.273*** | 0.048 | 0.184*** | 0.020*** | 0.243*** | -0.014 | 0.182*** | 0.015*** |
| | (0.038) | (0.043) | (0.004) | (0.004) | (0.045) | (0.053) | (0.005) | (0.004) |
| Unacceptable | -0.214*** | -0.004 | -0.303*** | -0.041*** | -0.168*** | 0.028 | -0.285*** | -0.019* |
| | (0.049) | (0.059) | (0.013) | (0.010) | (0.049) | (0.055) | (0.013) | (0.010) |
| Campus FE | No | Yes | No | Yes | No | Yes | No | Yes |
| Mean | -0.249 | | -0.092 | | -0.266 | | -0.095 | |
| Std Dev | 0.578 | | 0.449 | | 0.578 | | 0.450 | |
| N | 1406 | | 75891 | | 1143 | | 68501 | |

*Notes:* All regressions are at the campus-level of 1 year- and 2 year mean incoming student characteristics on indicators on school accountability rating dummies (excluded rating group is 'Acceptable'). All specifications include controls for time varying school characteristics and year indicators. Even numbered specifications include campus fixed effects. Standard errors in parentheses are clustered at the campus level. *, **, and *** denote significance at the 10%, 5%, and 1% levels respectively.