DOES IT PAY TO GET AN A?
SCHOOL RESOURCE ALLOCATIONS IN RESPONSE TO ACCOUNTABILITY RATINGS

Steven G. Craig[1]

Scott A. Imberman

and

Adam Perdue

University of Houston

October, 2009

Abstract

This paper examines whether school districts, and individual schools, respond to ratings from the accountability system by reallocating resources across or within schools. Our empirical work follows three identification strategies, a regression discontinuity for schools on the rating boundaries, a "rating shock" analysis for schools that face a change in rating when the state changed its accountability system, and a school fixed effects strategy. We find that school districts provided incentives for their schools to achieve higher ratings under the early accountability system, but under the later system they appear to have abandoned this strategy. In addition, the rating shock results suggest that some effort was directed towards assisting lower performing schools under the new regime. Finally, we find that in the early period incremental funds were used as much for ancillary purposes as instruction.

---

[1] Department of Economics, 204 McElhinney Hall, Houston, TX 77204-5019. The authors are grateful for helpful advice and comments from Aimee Chin, Meltem Daysal, Janet Kohlhase, Elaine Liu, Chinhui Juhn, and Lori Taylor, as well as participants at the University of Houston/Rice University applied microeconomics seminar, the meetings of the Southern Economics Association, and the American Education Finance Association. Correspondence should be sent to Scott Imberman at simberman@uh.edu or Steven Craig at scraig@uh.edu.

1. Introduction

Accountability systems have been fixtures of the US education system since the late 1990's. These systems, perhaps due to the pressures that arose out of state-wide equity standards, attempt to impose and publicize performance standards on schools. The primary role of these systems is to evaluate schools and assign simple ratings based primarily on student performance on statewide standardized tests. The ratings are clearly designed to be useful to parents. Figlio and Lucas (2004) find, for example, that there is a housing market response to the ratings in addition to measured learning output of schools, although this response seems to decline over time. Certainly if parents respond to school ratings, it is expected that school administrators would also use the ratings for allocation decisions. Even if parents do not respond to the ratings, however, school administrators may respond to other public pressures over actual or perceived school performance, and thus use the ratings to make resource allocation decisions. The goal of our research is to discern whether school administrators in fact use accountability ratings, by adjusting the allocation of resources in response. Our examination takes place on two levels. We first examine whether school districts change the allocation of resources between schools, by examining schools on the rating boundary. We additionally examine a wide variety of categories of spending and employee resources within schools. We find that in the early Texas accountability system school districts provided incentives to reward successful schools, but in the later reformed system the only response is to provide extra resources to schools with lower ratings. Similarly, we find that schools during the early period spread the incremental resources across categories of expenditure, which may include instructional spending, but particularly

increased spending on extra-curricular and athletic activities. Conversely, in the later period, incremental resources for lower rated schools were directed towards instruction.

Given that many researchers have found that schools have responded to accountability systems, it would be consistent to find resource allocation consequences. Incentives in some states, for example, are to support poorly performing schools. In other states the ratings serve only to provide easily accessible information to the community.[2] Economists have studied the extent to which schools respond to accountability systems by trying to increase test scores, and recent work has shown that schools do achieve this goal (Chakrabarti, 2007, 2008; Carnoy and Loeb, 2003; Hanushek and Raymond, 2004, 2005; Jacob, 2005). Researchers have also found, however, that some of these gains may be due to schools "gaming" the system (Cullen and Reback, 2006; Figlio, 2006; Figlio and Getzler, 2002; Figlio and Winicki, 2005; Jacob, 2005) or focusing on marginal students (Chakrabarti, 2007; Neal and Schazenbach, forthcoming).

Thus, while we know that schools respond to accountability in some finely detailed ways, we know little about the resource allocation response between or within schools. Understanding the resource response provides an important view of the incentives inherent in accountability systems, and also highlights that reallocations can potentially affect all students within a school, not simply those marginal students that affect a school's rating. While Rouse, Hannaway, Goldhaber, and Figlio (2007) provide evidence from a survey of schools that identify a number of policy changes that are induced by low ratings, there is very little evidence on how schools and districts re-allocate resources when they get lower ratings. Bacolod, Dinardo and Jacobson (2009) find that schools that receive rewards for higher ratings generally put the money into

teacher bonuses. Jacob (2003) looks at how school resources in Chicago adjust to the imposition of an accountability system, and finds shifts in expenditures to non-ancillary instruction amongst schools with low pre-accountability test scores but overall, he finds little change. Neither of these papers, however, look at resource allocation responses directly to the attainment of ratings. The one exception is Chiang (2009) who finds some evidence that schools who receive a "failing" grade in Florida increase spending on instruction and instructional tools. Nonetheless he only considers elementary schools for a single year. In this paper, we take much broader look at resource allocation responses to accountability - considering the response under two different accountability regimes and along different points of the rating distribution via multiple identification strategies.

Our investigation into the resource allocation response focuses on schools on rating boundaries. Accountability standards in Texas consist of clear demarcation based on the pass rates of students on a standardized exam administered by the state, along with attendance, dropout, and school completion rates. There are four ratings, exemplary (E), recognized (R), acceptable (A), and low performing (L).[3] Higher scores by one student do not compensate for lower scores by another; the criteria are solely based on whether a student scores higher than a minimum on the exam.[4] We thus use fluctuations in school ratings based on these criteria in three ways. First, we employ a regression discontinuity (RD) framework to account for schools on the boundary between each of the four ratings. Since there are random factors out of schools'

[2] Prior to recent court cases, students in Florida were eligible for vouchers to attend private schools providing a potentially strong punishment for poorly performing schools. On the other hand, in Texas punishments such as a school reconstitution are very rarely imposed making the system almost purely reputational.

[3] The "Low Performing" rating under the early accountability system was renamed to "Academically Unacceptable" under the later system.

[4] While they do not factor into the accountability ratings, schools with large percentages of students scoring at the "commended" performance level receive additional recognition.

control which partially determine student outcomes, schools that just barely receive a higher rating should be a valid comparison group for schools that just barely receive a lower rating (Lee, 2008). Our second analysis method exploits the "rating shock" that resulted when Texas implemented a more rigorous rating method in the middle of our period of analysis (Figlio and Kenny, 2009). The rating shock method examines budgetary changes in schools where the new system results in a decline in the rating. Third, we provide an analysis that identifies the effects of ratings on budget allocation using school fixed-effects. While this strategy allows a more general context since it uses all schools, it relies on the considerably stronger assumption that school ratings are uncorrelated with time-varying unobservable characteristics. Nonetheless, this strategy provides broadly similar results to our other two strategies.

Our findings show that the school districts in Texas provided additional resources to higher rated schools as an incentive device in the early period from 1997-2002. However, this strategy was abandoned when the state switched accountability systems starting in the 2003-04 school year. Further, we find some evidence that the school districts even reversed course under the new system and started to provide assistance to lower rated schools.

2. Texas Accountability Systems

Texas initiated one of the first educational accountability systems in 1993, called the Texas Assessment of Academic Skills (TAAS). Under TAAS schools were given ratings - from highest to lowest - of Exemplary (E), Recognized, (R), Acceptable (A), and Low Performing (L).[5] Table 1 shows the distribution of ratings by year. Very few schools receive an L rating

[5] In both of the accountability systems, alternative schools have the option of being evaluated under a separate system. Some schools with small populations are also rated using a subjective procedure called "special analysis."

under TAAS, and overall the number of schools receiving R and E ratings rose until the end of TAAS in 2002. After a one year transition, the new accountability system called Texas Assessment of Knowledge and Skills (TAKS) was implemented for the 2003-04 year. Under TAKS, the ratings distribution shifted downwards as fewer schools were awarded E's and more were rated L or A.

A school's accountability rating under TAAS is based on the share of tested students who pass the state-wide exam based on separate student groups in each subject. The groups are all tested students along with four subgroups - white, African-American, Hispanic, and economically disadvantaged. The subjects are math, reading, writing, and social studies (only for 8th grade). Dropout rates and attendance could also affect the rating.[6] Thus the system is based on test score levels rather than student gains, and the rating is determined by the lowest performing subject-group of sufficient size. Appendix Table 1 provides a description of the requirements for achieving each rating in the years of our analysis.[7] With the exception of receiving a low rating, there were no direct punishments imposed on schools by the state, and state awards for high performance were extremely small.[8] Schools that received an L were subject to additional oversight and students were given the right to transfer to other schools, although state law did not impose financial consequences. There was a risk of closure for being

---

Further some small schools were rated based on the performance of a nearby "paired" school rather than their own. We drop schools identified by the Texas Education Agency as being subject to one of these provisions from the analysis.

[6] The attendance requirements were abandoned in 1999-2000.

[7] More detail on the rules underlying the Texas accountability system can be found via the Texas Education Agency at http://ritter.tea.state.tx.us/perfreport/account/.

[8] Schools receiving an E or R rating, along with some A schools that made large gains, were eligible for financial awards. In 2000-2001, which was the last year the award system was fully funded, the award was provided at a rate of $7.20 per enrolled student up to a maximum award of $5000 per school. Average per-student expenditure that year amongst rated schools was $5490. Hence these awards too small to have more than a negligible effect on our estimates.

rated L for two or more consecutive years, although this affected very few schools as only 0.3% of all schools received an L rating two years in a row from 1998-99 to 2001-02. Thus, the rating system primarily acted as a reputation based system based on the public response, unless school districts themselves developed internal penalties or rewards based on the rating. Table 2 shows the transition matrix averaged over the years of TAAS, and shows schools often change ratings from year to year with only 69% of A's, 54% of R's, and 71% of E's maintaining their ratings the following year. On average ratings fall for 13% - 16% of schools and rise for 19% - 25% of schools. This upward drift over time contributed to implementation of a new accountability system starting in 2003.

The 2002-03 school year served as a transition period between the old system and the new system based on the TAKS exam. Under TAKS schools faced a similar rating system with four ratings - Exemplary (E), Recognized (R), Academically Acceptable (A), and Academically Unacceptable (L). Table 1 shows that the likelihood of an L rating, while still very low, triples compared to the earlier TAAS period while the likelihood of an E rating falls dramatically. Table 2 similarly shows that maintaining an E rating in TAKS is much less likely, while repeating an L rating is much more likely than under TAAS.

The TAKS system is structured similarly to TAAS but with a more difficult exam, some additional requirements for special education students and completions, and stricter passing requirements. Details on the requirements are provided in Appendix Table 2. As a result of the increased stringency, many schools experienced a drop in their rating after the transition, a situation we will exploit in our analysis. Punishments for low performing schools under TAKS

were strengthened, and include an option to "reconstitute" a school via mass layoffs and rehiring if a school receives an L for two consecutive years.

3. Empirical Methodology

Our examination of school district behavior is through the total operating expenses made available to each school. That is, the total budget available to each school in the district presumably reflects the school district's allocation strategy. On the other hand, we view the allocation of funds among categories of expenditure within each school as a reduced form potentially reflecting the preferences of both the school district, and the individual school principals.[9] We use all three of our empirical strategies on both aspects of potential resource reallocation due to ratings. Our RD strategy is segmented by each of the separate accountability systems, and is focused on schools near the rating borders. The rating shock strategy exploits the potential drop in rating facing schools as a result of the new exam in TAKS, compared to TAAS. Our final strategy uses all of the available data, but relies on school fixed effects to separate the impact of ratings on resources.

3.1. Regression Discontinuity Strategy

The key to the RD strategy is to define the margin that influences the response to the accountability rating. We believe that costs are most likely to influence school decisions, and thus we use the number of students for whom a change in test score could modify the school's

[9] An important related question is whether the test score gains represent actual improvements in learning, or are simply a result of other strategies. It may be that educational administrators' view of the answers to this question is reflected in the extent to which they view accountability as a useful piece of information for making resource allocation decisions.

rating. We further restrict our definition of boundary to identify schools which require improvement by test takers in a single subject/student group cell. With this definition, schools with the same distance from the boundary on one side should face similar costs of adjustment. For dropouts and completions we use the number of students in a single subject-group cell that would have to pass the exam, but did not, for a school to rise one rating (N_b). For dropouts and completions we use the number of students in each subject-group cell that need to stay in school to change the rating. The measure of N_b is therefore:

$$(1)\quad N_b = \sum_s \sum_g 1\left(Rating_{sg} = Rating\right) \times 1\left(Size_{sg} \geq Min_{sg}\right) \times N_{sg}$$

where *s* is the test subject or performance measure, *g* is the student group (by race, disadvantaged, or total), *Rating* is the school's accountability rating, $Rating_{sg}$ is the rating for the subject-group, and N_{sg} is the number of additional students in group g who need to pass performance measure *s* to achieve the next higher rating. The first indicator function limits the summation to subject-groups that, if they were to be rated individually, would have the same rating as the school as a whole.[10] The second indicator function limits the summation to subject-cutoff we measure N_a as

$$(2)\quad N_a = \min_{s,g} \left[1\left(Rating_{sg} = Rating\right) \times 1\left(Size_{sg} \geq Min_{sg}\right) \times N_{sg}\right]$$

[10] Since the overall school rating is based on the lowest rated subject-group, no group that counts towards the rating would have a rating lower than the school-wide rating. Nonetheless, under TAKS schools are given a limited number of exceptions for groups that perform poorly. Since these subject-groups do not count towards the ratings we do not include them in our calculation of N_b.

where N_{sg} refers to the number of students in a subject-group cell who would need to switch from passing to failing the performance measure to reach the cutoff for the next lower rating. As before, the two indicator functions identify subject group cells that have ratings equal to the school's rating, and that meet the minimum size standards.[11]

Thus we conduct a regression-discontinuity of schools whose N_a or N_b are close to zero for each boundary. Specifically, we conduct local-linear regressions of the form

$$(3)\quad R_{i,t+1} = \alpha + \beta_1 N_{b,it} + \beta_2 N_{a,it} + \beta_3 Above_{it} + \varepsilon_{it}$$

for school i in year t where $R_{i,t+1}$ is the school specific resource in category i (either dollars per student or inputs per student) the year after the school receives a rating, N_a and N_b as defined by (1) and (2) above, and *Above* is a dummy variable indicating whether the school is above the rating threshold. We use bandwidths selected through leave-one-out cross validation.[12] As suggested in Lee and Lemieux (2009), we use a rectangular kernel that involves limiting the sample to narrow bands around the cutoffs without re-weighting the data.[13] To explore whether schools and districts may react differently to receiving different ratings, we run (3) separately for

[11] In some years a school could achieve a higher rating using year-on-year increases in performance measures if they do not score high enough to meet the requirements, called required improvement (RI). In these cases we calculate N_b using whichever method, RI or standard, that brings that subject-group closer to the cutoff. Similarly, we use the RI calculation for Na if a school achieves a higher rating due to RI in the marginal group.

[12] Specifically, we select the margin that minimizes the mean squared error by repeatedly estimating the model for "all but one" of the observations with a wide variety of margins. Nonetheless, we have also estimated (3) using parametric techniques with a 5-order and a 3-order polynomial, and using bandwidths one unit higher and one unit lower than the cross-validation bandwidths. In all of these cases we find qualitatively similar results suggesting that our estimates are robust to the choice of bandwidths or functional form. These results are provided in the online appendix.

[13] Lee and Lemieux (2009) argue that more complex kernels provide only marginal improvements in efficiency.

each boundary - between L and A schools (the LA boundary), between A and R schools (the AR boundary), and between R and E schools (the RE boundary). Our results are based on estimates of β_3, the impact of being above the given boundary. Since the RD estimates result from marginal (random) changes in rating outcomes, the resulting estimate illuminates the response by either schools or school districts to a change in information revealed to the public, and not necessarily the response to an actual or perceived change in school quality. Figure 1 shows the first stage after the adjustments described above. At each boundary, there is a clear discontinuity in the rating whereby almost 100% of schools with $N_a \geq 0$ and $N_b = 0$ get a higher rating, with the near opposite occurring on the other side of the boundary.[14]

3.2. Rating Shock Strategy

As an alternative strategy to determine the budgetary response to school accountability ratings, we examine the transition period between the first accountability system in Texas, TAAS, to the newer system, TAKS. In particular, we test how schools that faced a rating reduction reallocated their budget for the first year in which the new TAKS system was implemented. This is similar to the strategy used by Figlio and Kenny (2009) due to a change in regimes. The TAKS system is designed to be more rigorous than TAAS in two dimensions. First, the exam given to students is intended to be more challenging and broader across academic material. Second, additional categories of evaluation for completions and test performance of special education students were added. Our empirical examination is based on the transition year between the two systems, as schools gave the new TAKS test in the year after TAAS, but the

[14] Given these results, we rely on a "strict" RD design. Nonetheless estimates using a "fuzzy" design were nearly identical and are available in the online appendix.

results were not publicized.[15] As a result, schools received good information about their performance and hence their new rating, and had the opportunity to alter their behavior in response. Since the TAKS exam was given for the first time, the information provided by the test results (the "rating shock") is plausibly exogenous and thus our estimates of the budgetary response are more likely to be unbiased. Controlling for the rating under the older TAAS system, we look for resource reallocations if a school projects using data from the transition year of 2002-03 that its rating will fall in the new system.[16] Like the RD results, our test of the effect from the shock of potentially lower accountability ratings is a test of the response of the school to a change in information provided to the community, rather than a change in actual output of the school. Unlike the RD strategy, however, since schools were not given an official rating, it is unlikely that any award system would be in place during this period. A further advantage of the rating shock strategy is that the RD provides a local average treatment effect only on the impact of ratings for marginal schools, while the rating shock strategy addresses a somewhat broader array of schools.[17]

The regression we run using the rating shock in the transition is therefore the change in total expenditures, categorical expenditures, or staffing from the 2002-03 school year to the 2003-04 school year based on whether the school projected that its accountability rating would

[15] Results of the testing were later provided as data releases which allow us to approximate what the schools' ratings would have been. There were no school report cards released, however, for the general public to evaluate schools.
[16] There was also uncertainty because the new criteria for translating pass rates to accountability ratings were not yet known. Nonetheless, we assume that school districts were able to approximate the new requirements prior to finalizing their expenditure decisions; hence we use the new (2003-04) rules to estimate the 2002-03 ratings.
[17] Below we provide a test for whether "marginal" schools are treated differently than infra-marginal schools. This test suggests the RD analysis captures the behavioral responses of schools and districts to ratings.

drop from its rating in 2001-02, the last year of the TAAS system.[18] Further, while controlling for the 2001-02 TAAS rating we interact an indicator for whether the rating falls with the rating the school is predicted to receive, so that we can identify whether falling to a lower predicted rating depends on the final projected rating of the school. Thus we estimate:

$$(4)\quad \Delta R_{i,2003-04} = \alpha + \beta_1 Fall_{i,2002-03} \times Rating\_L_{2002-03} + \beta_2 Fall_{i,2002-03} \times Rating\_A_{2002-03} + \beta_3 Fall_{i,2002-03} \times Rating\_R_{2002-03} + \beta_4 Rise_{i,2002-03} + \gamma_1 Rating\_L_{2001-02} + \gamma_2 Rating\_R_{2001-02} + \gamma_3 Rating\_E_{2001-02} + X_i\Omega + \varepsilon_i$$

where ΔR is the change in resources per student in a given category from the gap year (2002-03) to the next, either expenditures or employees for school *i*. X is a set of school characteristics including the percent of enrollment in each grade, percent of enrollment by racial category, and the percent of enrollment identified as economically disadvantaged. *Rating_L, Rating_R, and Rating_E* are indicator variables for whether school *i* received a rating of L, R, or E in the subscripted year. Our measure of a rating shock comes from the interaction of the predicted ratings with *Fall* - a dummy variable which equals one if scores during the gap year (2002-03) on the new test indicate the new TAKS rating will be lower than the last TAAS rating. We also control for a rise in rating, so that we compare schools whose ratings fall to those whose ratings remain the same. Hence our estimates compare schools that had the equivalent ratings in 2001-02 but where one received the same predicted rating in 2002-03 while the other received a lower rating. Other than for L schools, which account for only 1% of our sample for this analysis, schools were much more likely to experience a drop rather than increase in their rating. Table 1

[18] Due to the new testing regime, applying the "required improvement" rules to the predicted ratings would be not be sensible as the passing rates on TAKS were considerably different from those under TAAS. Thus, we ignore

shows that while only 16% of schools saw a ratings drop in the last year of TAAS, the rating from the first year of the TAKS exam suggests roughly two-thirds of schools would experience reduced ratings, while less than 3% of schools were predicted to receive a higher rating.

3.3. School Fixed Effects Estimation

As a final test of whether school accountability ratings affect resource allocations between and within schools, we estimate panel regressions from 1998-99 to 2006-07 separately for the two testing regimes with fixed effects for years and schools. As in the ratings shock analysis, we condition on the share of school enrollment in each grade, the racial and ethnic composition of the school, and the share of the students that are economically disadvantaged. Unlike the other methods, this strategy has the advantage of capturing the combined effect on resource allocation of both the signal provided by the rating to parents and the actual school quality underlying the rating. Nonetheless, the reliance on school fixed effects to identify the model implies that the estimates are unbiased only if there is no variation in omitted variables over time. Despite the dynamism of school populations, it is possible the included student characteristics successfully capture the relevant changes, leaving the core estimates of the impact on school accountability ratings informative of actual behavior of school administrators.

4. Data

Our data covers all public schools in the state of Texas, and comes from three datasets provided by the Texas Education Agency (TEA). First is the Academic Excellence Indicator System (AEIS) which provides data on staffing, enrollment, and student demographics. Second

---

required improvement here.

is the Public Education Information Management System (PEIMS) financial reports which provide expenditure data by category.[19] We examine total expenditures, four functions, five programs, and hiring in three categories of employees. Third is the TEA accountability reports which provides the data that is used to calculate accountability ratings.[20]

The PEIMS resource data contains school-level expenditures classified by functions, and programs. The four functions are aggregated into instruction; leadership, curriculum and staff development; counseling and social work services; and extra-curricular activities. The five programs are basic instruction; special interventions (called accelerated and alternative education by TEA); gifted and talented; career and technology; and athletics.[21] The AEIS data provides full-time equivalent staffing counts broken into teachers, aides, and administration. We normalize all of the variables by enrollment.

The RD regressions, as well as the fixed effects regressions, are pooled across years within each accountability system. Since some schools appear multiple times, we cluster standard errors by school. In all our analyses we drop alternative schools, which generally are targeted to specific groups of students, and which operate under a separate accountability system. Similarly, we drop charter schools, since they also are targeted at specific groups of students and have separate state aid and budgeting rules. We also drop schools that do not receive their own rating due to being "paired" with another school, schools that are identified as undergoing "special analysis" which is a subjective rating system for small schools, and schools which have

[19] We use the actual expenditures by schools rather than budgeted expenditures.
[20] All data sources are publically available on the website for the Texas Education Agency.
[21] Special interventions primarily include assistance provided to low-performing and at-risk students. The definitions for programs changed slightly starting in the 2002 school year. Fortunately, this coincides with the change in the accountability systems which we estimate separately, so there is no impact on the empirical work.

fewer than 200 students, to avoid errors in calculating margins due to some masked data.[22] After these restrictions, we have 26,500 school-year observations for TAAS and 23,168 for TAKS.

Table 3 presents the means of school characteristic data for each of the two accountability regimes. Schools with higher ratings tend to have fewer minorities, disadvantaged, LEP, and special education students as well as smaller enrollments. Table 4 provides summary statistics for resources in the year after a school receives a rating. Under both accountability regimes schools with lower ratings have higher expenditures than those with higher ratings, although this gap widens in TAKS.

5. Empirical Results

5A. Regression Discontinuity

Table 5 presents the RD results for school district behavior for three alternative bandwidths, by illustrating how school districts alter total spending per student for schools that are just above a rating border, compared to schools that are just below. The results in panel A, for TAAS, show estimates that are consistently positive, indicating that school districts reward schools that are just above the border on all three margins. The magnitude of the estimates suggests a considerably larger reward for A compared to L schools than on the other two borders. The point estimate on the LA border of $347 is about 6.8% of the average spending by A rated schools ($5,075), although it is not significant at conventional levels. Small variation in the bandwidths produce estimates that are larger in magnitude with similar standard errors, hence we interpret the bulk of the evidence as indicating rewards are produced by school districts for

[22] Student counts in the accountability reports are masked when sample sizes are too small, generally fewer than 5 students.

schools that successfully jump their specific rating boundary. The pattern of the LA border is repeated for the AR border, and the estimates are consistently positive for the RE border. Thus while the rewards for success are much smaller at the two higher boundaries, the evidence shows a consistent pattern in TAAS of school districts providing rewards to successful schools as indicated by the school rating.

Figure 2 provides a graphical representation of these results. The figure provides the average expenditure level for the number of students below and above the boundary along with a fitted polynomial from local linear regressions. The figures generally support the empirical results, in that the discontinuity is much more pronounced for the LA and AR borders, while the RE border results disappear rapidly with distance from the border.

In panel B of Table 5 we see that the implicit rewards in TAAS for success at the border have completely disappeared for TAKS. To the extent school districts use the accountability rating results to affect resource allocations in TAAS, they no longer do so with the newer accountability system. Figure 3 graphically presents the same information - that there are no evident discontinuities on the boundaries. We believe it unlikely that the stricter environment in TAKS is responsible for the change, in that while L schools faced a larger administrative penalty, other schools did not. Thus it seems more likely that school districts abandoned the implicit reward structure because they found it was not successful, although our estimates cannot distinguish the actual cause of the change in behavior.[23]

[23] To the extent school districts represent derived demand, this implies parents did not find the ratings useful.

Table 6 provides estimates of how ratings affect different expenditure and staffing categories under TAAS.[24] In general, the impact of higher ratings on each individual expenditure category is relatively small, suggesting that schools dispersed the funding they were awarded from TAAS success across a variety of categories. We find significant increases in administration & training, as well as counseling, on the A/R and R/E boundary, and we find increases in gifted and talented spending on the A/R border. Nonetheless all of these are economically small, amounting to a total of $57 per-student for E schools and $48 per-student in R schools. A clearer story emerges when looking at the L/A boundary, although small sample sizes make the estimates rather imprecise. In this case it appears that a substantial portion of additional funds went into extra-curricular activities and instruction. "A" schools spent an additional $86 per-student on extra -curricular activities, a substantial increase over the $99 average. Figure 4 highlights this increase - there is a clear jump in extra-curricular spending at the L/A boundary. Such a response is consistent with schools increasing their expenditures on categories unrelated to test-preparation as ratings increase and accountability pressures relax. Similarly expenditures on athletics are found to increase by $286 for highs schools, although the relatively small sample size warrants a cautious interpretation. Nonetheless, the extra spending on ancillaries is also accompanied for A schools by a substantial, although insignificant, increase in instruction expenditures of $347.[25]

Table 7 presents the categorical spending results for the later system, TAKS. For A schools along the L/A boundary, the only marginally significant result is a drop in extra-

[24] Regressions using bandwidths one above and one below the leave-one-out cross-validation optimal bandwidths were qualitatively similar and are provided in the online appendix.

curricular spending, the opposite of what we found for TAAS. For R schools along the A/R boundary there is an increase shown in counseling and G&T expenditures but a drop in athletics and student-teacher ratios increase (teachers fall). E schools are found to show a marginal increase in extra-curricular spending but also an increase in the student-aide ratio (aides fall). Hence, these results appear to confirm the theory that arose out of the total expenditure results that districts abandoned their reward systems under the TAKS regime, as we find no systematic return for being on the positive side of a rating boundary.

As a diagnostic test for the above results, Table 8 presents tests for discontinuities in the demographic variables, which ideally are randomly distributed around the rating cut-off. To conduct this test we use the optimal bandwidths for total expenditures.[26] In general we find there is no significant difference between schools above or below any boundary in % white or % disadvantaged. In addition, there appears to be no significant impact of the rating on the change in enrollment.

While this test generally supports the RD framework, there are a few cases nonetheless where there are estimated discontinuities. E schools near the R/E boundary in TAAS are found to have moderately higher enrollment than R schools. A similar result holds for R schools at the A/R boundary in TAKS. We note, however, that on average E schools in TAAS and R schools in TAKS have lower enrollment than R and A schools, respectively. Hence, at worst the R-D tends to overcompensate for enrollment differences along these boundaries. Additionally, the other four enrollment estimates are statistically insignificant. The other significant estimates are for %

[25] The one peculiar outcome is a large drop in career-tech expenditures of $500 per-student. However, as with athletics the sample size is very small and is sensitive to choice of bandwidth and hence this result may simply be spurious. Increasing the bandwidth by one reduces this figure to an insignificant -119.6(se 170.4).
[26] Results using the median bandwidth across all outcomes were similar.

black; R schools on the A/R boundary are found to have a higher share than A schools in TAAS and a lower share but more Hispanics in TAKS. These differences are small, however, as the difference in total in minority share does not exceed 1.9% in either case.

Another potential concern is that schools near a boundary may try to manipulate their accountability measures to get a higher rating. While such "gaming" likely occurs, this would not be a problem if at least some of the determinants of the distance from a ratings boundary, and by extension the rating itself, are random (Lee and Lemieux, 2009). In the online appendix we provide distributions for TAAS and TAKS, respectively. In both cases, the distributions seem to fit a smooth pattern except precisely at the boundary (N = 0) or just below it (N = -1), where N = 0 is more likely than a smooth graph would predict and N = -1 is less likely. This pattern is more apparent in TAKS than in TAAS. We believe the pattern potentially results from some manipulation, but is also a product of the Rawlsian nature of the rating algorithm. In TAKS, schools were allowed a limited number of exceptions whereby some subject/student groups that would have given the school an L rating were not counted. Indeed, omitting the exceptions considerably reduces the discontinuity at the border. Hence to address this concern, since any manipulation appears to be precisely concentrated at N = 0 and N = -1, we conduct analyses using the CV optimal bandwidths dropping these marginal observations from the regressions. These results can be found in the online appendix, where we show there are only two coefficients significantly different when the N = 0 and N = -1 observations are dropped, and neither of these results are for total expenditures in TAAS. Rather, one is for extra-curricular on the AR border in TAAS, and the other is for instruction on the AR border in TAKS. Thus we believe our central results are not sensitive to the potential "gaming."

We conduct a number of additional specification tests provided in the online appendix. We find results that are generally similar to our baseline estimates when we restrict all bandwidths to the optimal bandwidth for total expenditures, when we limit the sample below a boundary to schools with one failing subject instead of a single failing subject-group, and when we use a 3 term or a 5 term polynomial instead of local-linear regressions. In addition, a small number of schools (less than 0.5%) are calculated to have ratings different from what their ratings actually were. This is because of masked data for small student-group cells, and an attendance category in the early years of TAAS that we were not able to incorporate into our rating calculation. Our main reported estimates minimize the effects of these attributes by dropping schools with less than 200 enrollees. Nonetheless, we further test our results by using the actual rating as the intercept term instead of our calculated rating, and use the calculated rating as an instrument. Results for this analysis are also similar to our baseline results.

Our final set of sensitivity analyses explores dividing the sample. Despite the relatively small numbers of middle and high schools, we split the sample by elementary and the aggregate of middle and high schools. This shows generally similar results across elementary and middle/high schools, with the exception of L/A in TAAS where most of the expenditure increase for A schools is found in middle/high schools, as is most of the increase in extra-curricular expenditures. These schools also experience a reduction in student-teacher ratios. Hence, most of the benefits here appear to accrue to middle/high schools although elementary schools show positive but insignificant estimates on instruction and regular program spending. The caution in these results is because despite pooling the middle and high schools, the sample sizes are much smaller than the overall results. Finally, we split our sample also by the large urban districts

(Houston, Dallas, San Antonio, Austin, Fort Worth, & El Paso) and other districts. Results were qualitatively similar for both types of districts.

5b. Ratings Shock Results

Our second identification strategy is to use the potential unanticipated reduction in rating facing many schools from the transition between TAAS and TAKS, which we refer to as a "rating shock" as in Figlio and Kenny (2009). This rating shock arises because schools and districts did not know how they would be affected by the new exam until it was administered, and had a year to respond before the results were made public. Thus to perform the rating shock analysis we estimate the rating using test data from the gap year of 2002-03 and the thresholds actually adopted in 2003-04.[27]

The regression results in Table 9 show the change in resources from the 2002-03 transition year to the first official year of TAKS (2003-04) as a response to the reduction in the school rating, differentiated by the rating to which a school falls. The evidence suggests school districts responded to prop up schools which were threatened with a rating drop. Schools that were expected to fall to an A rating receive an average of $46 more per-student than otherwise. Most of this funding appears to go towards instruction ($45), which leads to a reduction in student-teacher ratios (teachers rise), and is also shown by the regular program results ($60). Notably, special interest students suffer a drop of $28 per student. For schools whose predicted rating falls to L, with the exception of total expenditures, the point estimates are in the same direction as those that fall to A for the categories listed above, albeit insignificant. Further, these

[27] We are thus assuming here that the schools and districts are able to anticipate the new rating criteria, and so could adjust their expenditures in response.

schools also experience an increase in student-teacher ratios.[28] Finally, for schools that fall to an R we find no statistically significant changes in any resource.

5c. School Fixed Effects

Our third identification strategy is to use the entire panel of data but rely on school fixed-effects to identify the impact of ratings. This strategy has the advantage of exploiting the entire data range rather than a sub-population of "marginal" schools. It has the substantial drawback, however, that it relies on stronger assumptions about the nature of unobserved covariates than the two procedures above.

Tables 10 and 11 provide the fixed effects estimates for TAAS and TAKS, respectively. The estimates are broadly consistent with our findings, in that districts apparently provide incentives to schools to increase their ratings under TAAS, and schools are found to use that money to fund ancillary programs. Under TAKS schools and districts instead provide additional funding to lower rated schools. For example, in Table 10 we see that under TAAS higher rated schools get a small and statistically insignificant amount of additional money overall but that there are significant, albeit small, increases in extra-curricular activities for higher rated schools. In Table 11 for TAKS, the estimates suggest that schools with L ratings receive more overall funding as well as more funding in administration and training, counseling, extra-curricular activities, and special interventions along with a lower student-teacher and student-administrator ratios. For other ratings the estimates are generally similar except for extra-curricular activities where A schools receive more funding than R and E schools.

[28] We find no statistically significant change in enrollment as a result of the rating shock, thus we interpret this as teachers changing.

5d. Marginal vs. Infra-marginal Schools

An alternative model is that schools and school districts respond to being on the margin of a rating, whereas which side of the border they are on is not relevant to the resource response. To test this idea we estimate whether being an A school that is close to either the L/A or A/R boundaries has a significant correlation with funding relative to being an A school that is far from either boundary. We also conduct a similar analysis for R schools. We consider being close to a boundary as being 3 or fewer students away, although results are similar using a margin of 5.[29] These results are provided in Table 12 for TAAS and Table 13 for TAKS. While there are some cases where one boundary has a significant estimate, in no case does being infra-marginal have a clear impact on resources relative to being marginal. Hence it appears that schools and districts respond to the actual rating rather than treating all schools on the margin equivalently compared to infra-marginal schools.

6. Summary and Conclusion

Our objective in this research has been to establish how schools and school districts reallocate resources in response to accountability ratings. Using data from the state of Texas, our RD analysis of school on the rating boundary (defined by number of students and a single subject group) has found two broad results. First, we find that school districts appear to have rewarded success based on accountability ratings during the early, TAAS, accountability system. A considerable portion of these extra funds appear to have been used for extra-curricular activities and athletics. The second result, however, is that schools and their districts apparently did not

find this incentive structure useful. When Texas switched to the more rigorous TAKS system, these incentive effects are no longer apparent. The rating shock analysis even shows that schools which had a drop in rating to A were given incremental resources to bolster their instructional programs.

It is unclear why the new accountability system would engender a different response than the old system. One possibility is that in an environment with fixed resources, funding for successful schools can only come at the expense of unsuccessful schools. Further, if incremental resources are used for non-instructional purposes, school administrators may feel such funds are mis-directed. Another possibility is that the increased emphasis in TAKS on avoiding low ratings altered the perceived incentives on the margin.[30] A final possibility, consistent with Figlio and Lucas (2004), is that schools and parents have learned over time that accountability ratings poor predictors of school quality, and thus the ratings do not provide useful information.

We believe these results are an important extension to the literature on accountability, as they expand the discussion to consider whether resources are being channeled to the most effective institutions. Sonstelie and Silva (1995) find that the original school finance equalization lawsuit in California lead to the evisceration of the property tax system through Proposition 13, so that local schools are almost entirely financed by state government. If a similar process holds for other states as equalization spreads, it seems natural to examine whether accountability assists in developing an allocation system to replace the Tiebout system of population movement. Our research here suggests that accountability systems are not, at least in Texas, replacing the Tiebout process of allocating resources to the most successful institutions.

[29] These are provided in the online appendix.
[30] This is also consistent with avoiding low ratings from the federal No Child Left Behind law.

Nonetheless, accountability systems require significant resources from both state governments and school administrators, and if it is a general result across states that they fail to generate allocation responses, it may be that a superior mechanism can be formulated to replace the Tiebout process.

References

Bacolod, Marigee, John DiNardo, and Mireille Jacobson, "Beyond Incentives: Do Schools Use Accountability Rewards Productively?" NBER Working Paper 14775, 2009

Carnoy, Martin and Susanna Loeb. "Does External Accountability Affect Student Outcomes? A Cross-State Analysis." *Education Evaluation and Policy Analysis*, 24, 2003, pp. 305 - 331.

Cullen, Julie Berry and Randall Reback. "Tinkering Toward Accolades: School Gaming under a Performance Accountability System." NBER Working Paper #12286, June 2006.

Chakrabarti, Rajashri. "Impact of Voucher Design on Public School Performance: Evidence from Florida and Milwaukee Voucher Programs." Federal Reserve Bank of New York Staff Report no. 315, January 2008.

---. "Vouchers, Public School Response and the Role of Incentives: Evidence from Florida." Federal Reserve Bank of New York Staff Report no. 306, October 2007.

Chiang, Hanley. "How Accountability Pressure on Failing Schools Affects Student Achievement." *Journal of Public Economics, 93, 2009, pp. 1045-1057.*

Figlio, David and Lawrence Kenny. "Public Sector Performance Measurement and Stakeholder Support." *Journal of Public Economics, 93, 2009, pp. 1069-1077.*

___and Cecelia Elana Rouse. "Do Accountability and Voucher Threats Improve Low-Performing Schools?" *Journal of Public Economics*, 90, 2006, pp. 239 - 255.

___ "Testing, Crime, and Punishment." *Journal of Public Economics*, 90, 2006, pp. 837 - 851.

___ and Lawrence Getzler, ""Accountability, Ability, and Disability: Gaming the System," in *Advances in Microeconomics*, Timothy Gronberg (ed), Elsevier, 2006.

___ and Joshua Winicki. "Food for Thought: The Effects of School Accountability Plans on School Nutrition." *Journal of Public Economics*, 89, 2005, pp. 381 - 394.

___ and Maurice E Lucas, "What's in a Grade? School Report Cards and the Housing Market," *American Economic Review*, 94, June, 2004, pp. 591-604.

Hahn, Jinyong, Petra Todd, and Wilbert Van Der Klaauw, "Identification and Estimation of Treatment Effects with a Regression-discontinuity Design," *Econometrica*, 69, January 2001, pp 201-09.

Hanushek, Eric A. and Margaret E. Raymond. "Does School Accountability Lead to Improved Student Performance?" *Journal of Policy Analysis and Management*, 24, 2005, pp. 297 - 327.

--- and ---. "The Effect of School Accountability Systems on the Level and Distribution of Student Achievement." *Journal of the European Economic Association*, 2, April-May 2004, pp. 406 - 415.

Jacob, Brian A. "Accountability, Incentives, and Behavior: The Impact of High-Stakes Testing in the Chicago Public Schools." *Journal of Public Economics*, 89, 2005, pp. 761 - 796.

---. "Getting Inside Accountability: Lessons from Chicago." *Brookings-Wharton Papers on Urban Affairs*, 2003, pp. 42 - 70.

Kane, Thomas J., Douglas O. Staiger and Gavin Samms. "School Accountability Ratings and Housing Values." Brookings-Wharton Papers on Urban Affairs, 2003, pp. 83 - 127.

Lee, David S., "Randomized Experiments from Non-Random Selection in U.S. House Elections," Journal of Econometrics, 142, 2008, pp. 675-97.

Lee, David S. and Thomas Lemieux. "Regression Discontinuity Designs in Economics." NBER Working Paper #14723, February 2009.

Neal, Derek and Diane Whitmore Schanzenbach. "Left Behind by Design: Proficiency Counts and Test-Based Accountability." Review of Economics and Statistics, forthcoming.

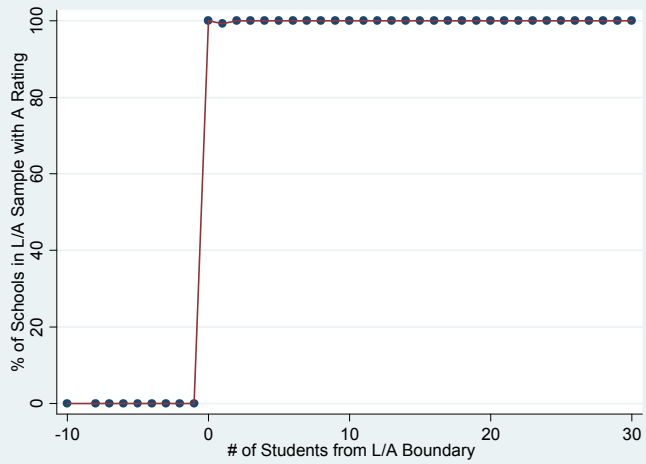
Rebeck, Randall. "Teaching to the Rating: School Accountability and the Distribution of Student Achievement." *Journal of Public Economics*, 92, 2008, pp. 1394 - 1415.

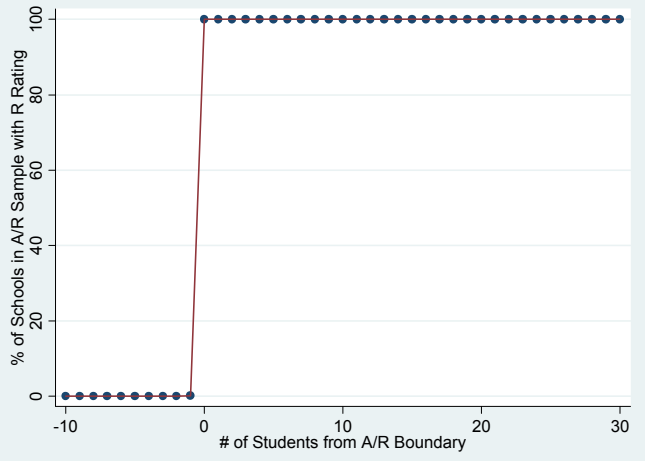
Rouse, Cecelia Elena, Jane Hannaway, Dan Goldhaber and David Figlio. "Feeling the Florida Heat? How Low-Performing Schools Respond to Voucher and Accountability Pressure." NBER Working Paper #13681, December 2007.

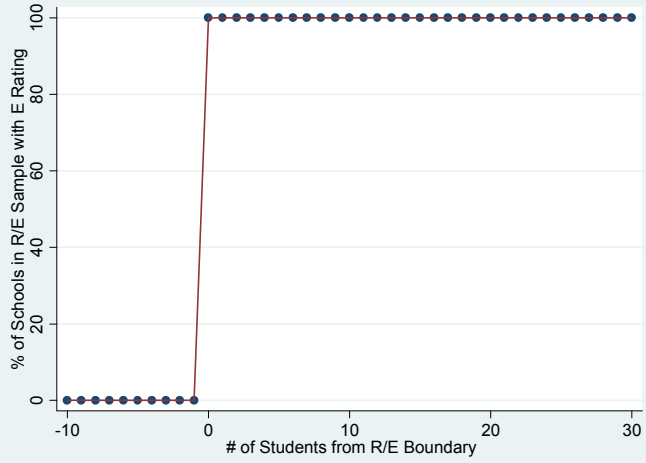
Silva, Fabio and Jon Sonstelie, "Did Serrano Cause a Decline in School Spending?", National Tax Journal, 48, June, 1995, 199-215.

Figure 1: Changes in Accountability Rating at Rating Boundaries

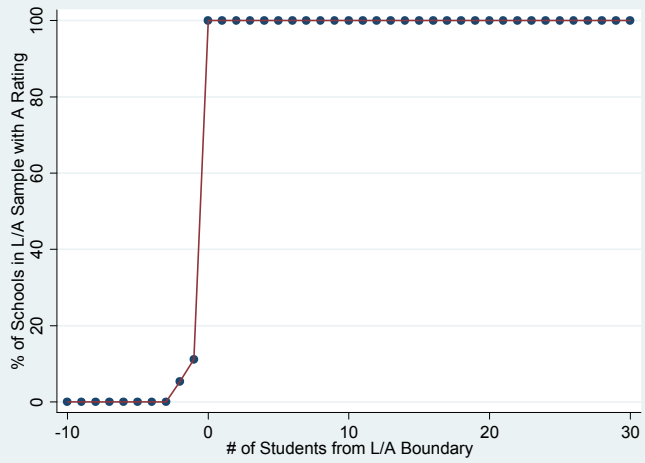
A. TAAS

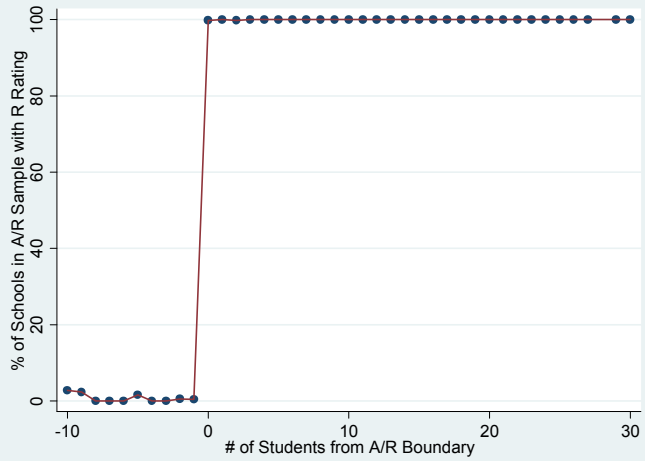


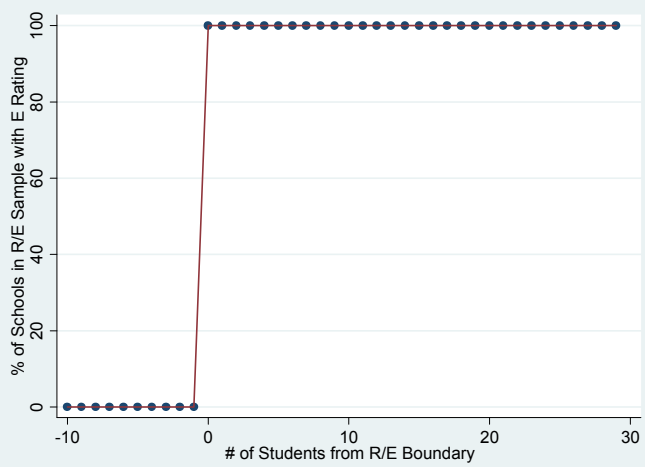




B. TAKS

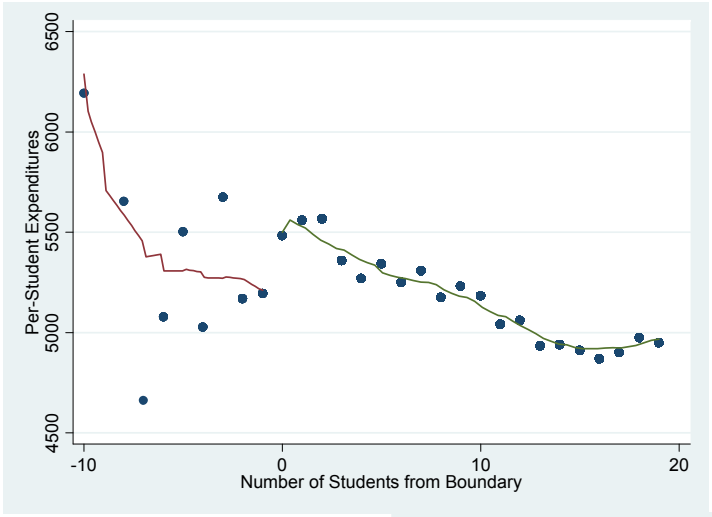




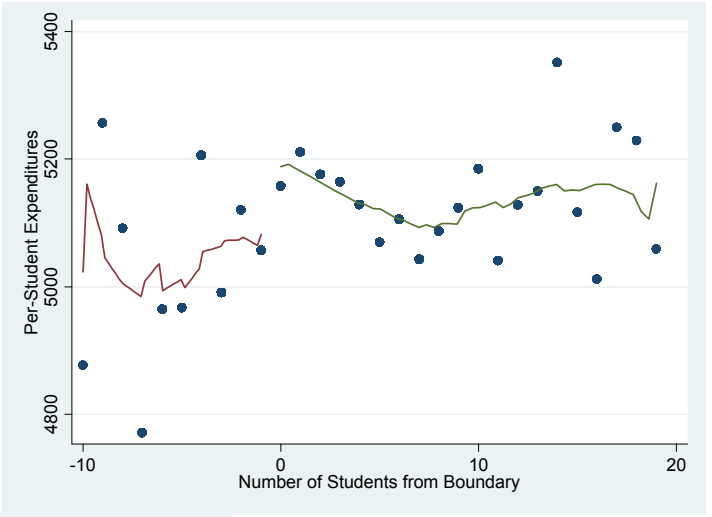


Sample is restricted to schools one rating above boundary or one rating below boundary where they miss the higher rating in only one subject-group. Schools that are paired with another school, received a rating on appeal, were identified as receiving "special analysis", or had fewer than 200 students are also excluded.

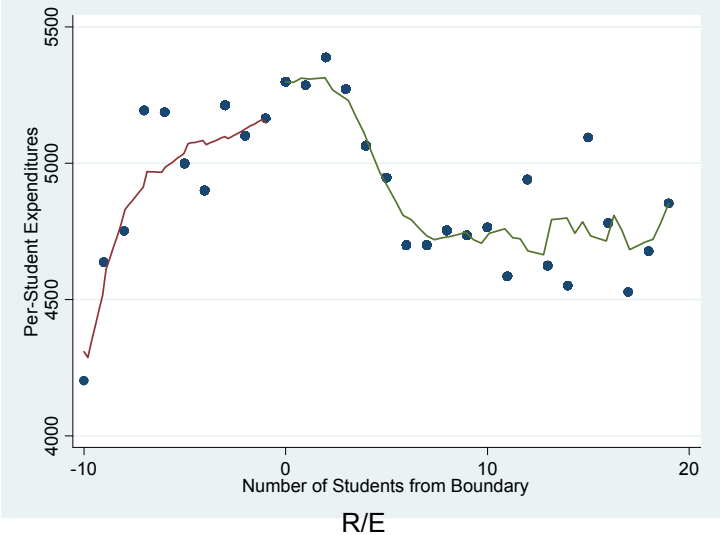
Figure 2 - Total Per-Student Expenditures in TAAS



L/A



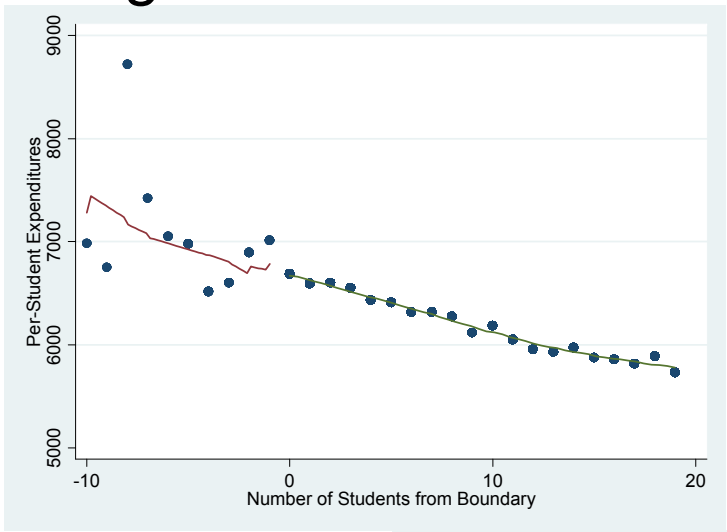
A/R



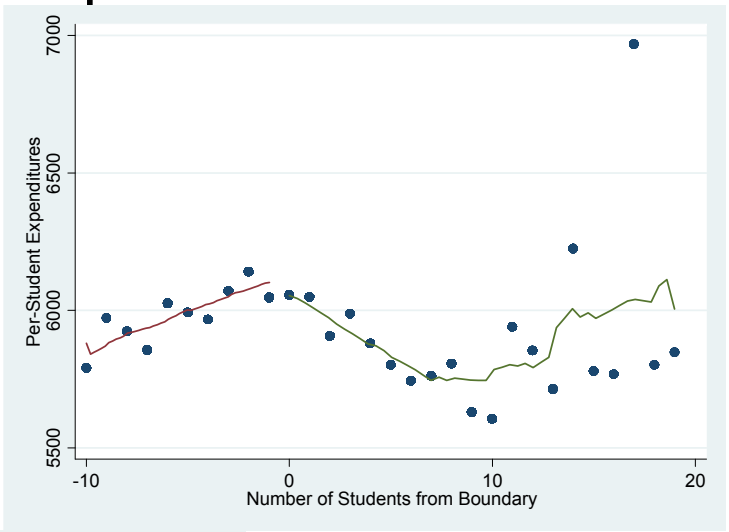
R/E

Sample is restricted to schools one rating above boundary or one rating below boundary where they miss the higher rating in only one subject-group. Schools that are paired with another school, received a rating on appeal, were identified as receiving "special analysis", or had fewer than 200 students are also excluded.

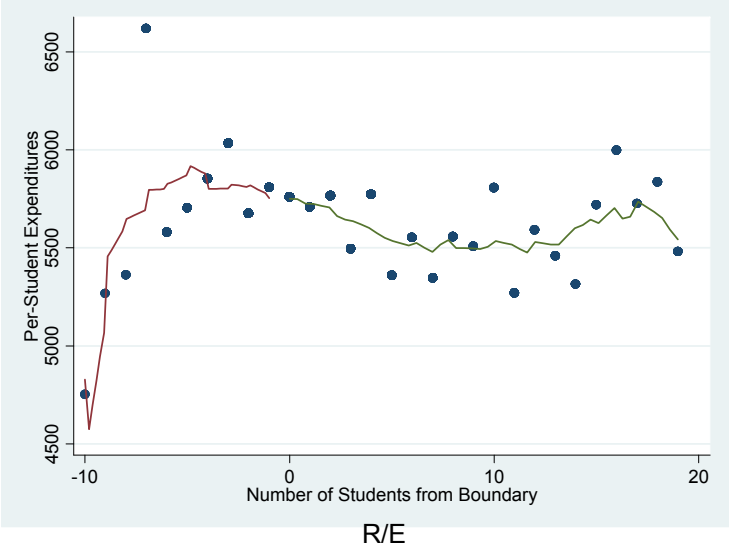
Figure 3 - Total Per-Student Expenditures in TAKS



L/A



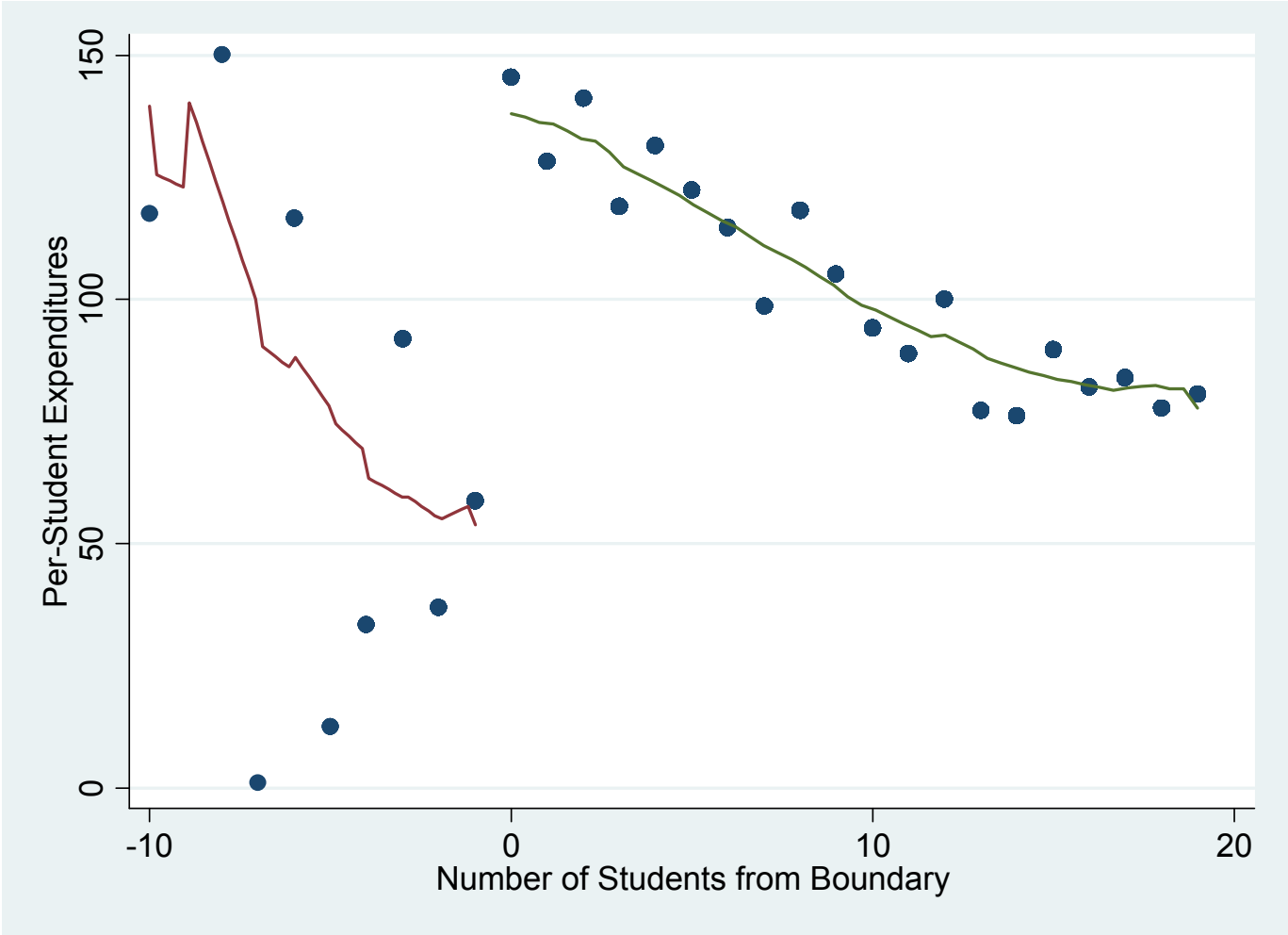
A/R



R/E

Sample is restricted to schools one rating above boundary or one rating below boundary where they miss the higher rating in only one subject-group. Schools that are paired with another school, received a rating on appeal, were identified as receiving "special analysis", or had fewer than 200 students are also excluded.

Figure 4: Extra and Co-Curricular Activites
TAAS, L/A Boundary



Sample is restricted to schools one rating above boundary or one rating below boundary where they miss the higher rating in only one subject-group. Schools that are paired with another school, received a rating on appeal, were identified as receiving "special analysis", or had fewer than 200 students are also excluded.

Table 1: Distributions of Accountability Ratings

	Accountability Rating				Rating Changes		
	Low	Acceptable	Recognized	Exemplary	% Lower	% Same	% Higher
TAAS							
1997-98	0.7%	55.1%	27.1%	17.2%	-	-	-
1998-99	0.9%	51.1%	29.9%	18.2%	14.9%	66.5%	18.6%
1999-00	1.6%	45.9%	32.0%	20.5%	14.8%	64.7%	20.5%
2000-01	0.7%	38.2%	36.5%	24.7%	13.2%	61.9%	24.9%
2001-02	1.3%	31.8%	37.4%	29.5%	16.4%	58.8%	24.8%
Gap Year (2002-03)							
Estimate based on 2003-04 Rules†	9.6%	69.7%	18.0%	2.7%	63.1%	33.4%	3.5%
TAKS							
2003-04‡	1.0%	53.4%	37.9%	7.7%	42.6%	39.8%	17.6%
2004-05	3.1%	64.3%	28.2%	4.4%	25.8%	66.9%	7.3%
2005-06	3.6%	46.6%	41.8%	8.1%	8.3%	64.2%	27.5%
2006-07	3.4%	53.6%	34.0%	9.1%	20.3%	65.0%	14.8%

† Only for schools that have a regular rating in 2001-02. We do not account for required improvement as the implementation of a new exam makes such a calculation inappropriate

‡ Ratings changes are relative to 2001-02.

Sample is limited to schools that received an L, A, R, or E rating. For changes the school must have received a rating in both years. Schools that received ratings on appeal, were paired with another school, were identified as having undergone "special analysis," or have fewer than 200 students are excluded (see text).

Table 2: Rating Transitions Under Different Accountabilty Regimes

A. TAAS

	Year t			
Year t - 1	L	A	R	E
L	11.3%	82.3%	6.4%	0.0%
A	2.0%	69.1%	25.5%	3.4%
R	0.1%	24.6%	53.6%	21.7%
E	0.0%	4.1%	24.6%	71.3%

B. Transition

Last Year of	First Year of TAKS (2003-04)			
TAAS (2001-02)	L	A	R	E
L	5.6%	80.6%	12.5%	1.4%
A	1.9%	75.9%	21.3%	1.0%
R	0.2%	59.3%	38.0%	2.5%
E	0.1%	29.2%	53.8%	17.0%

C. TAKS

	Year t			
Year t - 1	L	A	R	E
L	21.8%	73.3%	4.8%	0.3%
A	4.7%	74.0%	20.9%	0.3%
R	0.2%	33.8%	57.9%	8.1%
E	0.0%	3.1%	40.5%	54.4%

Sample is limited to schools that received an L, A, R, or E rating in both years. Schools that received ratings on appeal, were paired with another school, were identified as having undergone "special analysis," or have fewer than 200 students are excluded (see text).

Table 3 - School Characteristics by Rating

	TAAS				TAKS			
	L	A	R	E	L	A	R	E
% Asian	1.1	1.8	2.0	3.2	0.7	1.9	3.1	6.4
	(2.3)	(3.4)	(4.1)	(5.3)	(1.5)	(3.5)	(5.5)	(8.5)
% Black	29.5	18.6	11.1	7.1	30.2	15.6	11.0	7.3
	(26.8)	(22.6)	(16.6)	(11.7)	(29.8)	(19.4)	(14.7)	(10.2)
% Hispanic	48.4	43.8	39.4	24.5	52.6	49.3	39.6	22.0
	(30.1)	(31.7)	(32.3)	(28.6)	(32.3)	(31.5)	(30.9)	(24.7)
% White	20.8	35.6	47.1	64.9	16.4	32.8	45.9	63.9
	(23.6)	(29.0)	(31.3)	(29.1)	(22.4)	(29.0)	(30.4)	(26.1)
% Economically Disadvantaged	70.9	59.9	51.0	31.9	76.9	62.4	51.6	28.1
	(22.5)	(25.0)	(26.4)	(28.0)	(19.0)	(24.9)	(27.2)	(28.0)
% LEP	23.5	16.6	13.3	8.0	19.6	17.7	14.8	9.5
	(23.1)	(19.4)	(17.7)	(14.3)	(22.3)	(19.8)	(18.0)	(13.9)
%Special Ed	18.3	14.1	12.0	13.6	28.8	19.9	6.5	1.8
	(26.4)	(23.7)	(23.8)	(26.6)	(30.4)	(28.7)	(17.7)	(10.7)
Enrollment	812	749	633	601	781	783	606	595
	(503)	(502)	(410)	(383)	(519)	(567)	(336)	(226)
Observations	285	12,111	8,614	5,550	593	12,537	7,985	1,537

Sample is limited to schools that received an L, A, R, or E rating. Schools that received ratings on appeal, were paired with another school, were identified as having undergone "special analysis," or have fewer than 200 students are excluded (see text).

Table 4 - School Finances and Resoruces by Rating

Resources in Year t + 1	TAAS (year t)				TAKS (year t)			
	L	A	R	E	L	A	R	E
Function Expenditures								
Total Operating	5,520	5,075	5,152	5,159	7,023	6,219	5,926	5,674
	(1,734)	(1,093)	(1,114)	(1,341)	(1,570)	(1,221)	(1,231)	(1,000)
Instructional	3,980	3,728	3,843	3,863	4,936	4,492	4,433	4,341
	(1,463)	(748)	(740)	(880)	(961)	(797)	(830)	(687)
Admin & Training	587	498	485	475	770	613	566	532
	(267)	(181)	(210)	(298)	(278)	(178)	(177)	(148)
Counseling	215	196	186	182	286	239	213	198
	(109)	(91)	(85)	(89)	(127)	(108)	(98)	(90)
Extra-Curricular	98	99	110	127	212	177	78	35
	(162)	(178)	(211)	(246)	(314)	(288)	(184)	(87)
Program Expenditures								
Regular	2,888	2,866	3,005	3,197	3,585	3,402	3,492	3,649
	(799)	(653)	(687)	(711)	(923)	(822)	(878)	(715)
Gifted & Talented	63	72	78	91	66	79	91	108
	(119)	(151)	(154)	(172)	(162)	(163)	(203)	(240)
Career & Tech (High only)	611	666	730	761	787	722	747	757
	(269)	(273)	(297)	(353)	(398)	(308)	(423)	(628)
Special Interventions	842	581	531	363	1,061	721	650	400
	(1,275)	(437)	(495)	(381)	(752)	(548)	(599)	(482)
Athletics (High only)	225	285	381	404	330	410	417	259
	(169)	(219)	(247)	(282)	(265)	(311)	(343)	(301)
Student - Staff Ratios								
Student - Teacher	15.2	14.9	14.7	14.7	14.0	14.6	14.8	15.2
	(2.6)	(2.5)	(2.5)	(2.5)	(2.4)	(2.3)	(3.1)	(1.7)
Student -Aide	103.1	109	94	114	108.0	107	91	105
	(70.6)	(152)	(130)	(160)	(98.9)	(296)	(146)	(96)
Student - Administrator	305.8	325	331	335	251.9	300	315	338
	(125.0)	(168)	(198)	(166)	(106.3)	(340)	(143)	(112)

Sample is limited to schools that received an L, A, R, or E rating. Schools that received ratings on appeal, were paired with another school, were identified as having undergone "special analysis," or have fewer than 200 students are excluded (see text).

Table 5: Regression Discontinuity: The Effect of Rating on Per-Student Total Operating Expenditures in Following Year

	A. TAAS			B. TAKS		
	CV - Optimal Bandwidth Minus 1	CV - Optimal Bandwidth	CV - Optimal Bandwidth Plus 1	CV - Optimal Bandwidth Minus 1	CV - Optimal Bandwidth	CV - Optimal Bandwidth Plus 1
	(1)	(2)	(3)	(4)	(5)	(6)
Low / Acceptable	603.1**	346.9	440.5*	-230.2	-109.2	2.7
	(305.2)	(278.0)	(261.8)	(406.3)	(359.7)	(376.7)
Obs	336	525	785	1751	2353	3053
Bandwidth Below	3	4	5	5	6	7
Bandwidth Above	2	3	4	3	4	5
Acceptable / Recognized	187.9**	85.6	154.2**	9.4	-72.1	-71.4
	(95.1)	(68.8)	(72.1)	(96.4)	(80.1)	(69.4)
Obs	6880	8220	9086	4953	6021	6979
Bandwidth Below	2	3	4	3	4	5
Bandwidth Above	4	5	6	3	4	5
Recognized/Exemplary	70.0	154.3*	71.7	-182.0	85.8	54.2
	(118.6)	(93.4)	(82.6)	(242.2)	(202.6)	(156.8)
Obs	2237	3562	4535	1177	1463	1647
Bandwidth Below	2	3	4	2	3	4
Bandwidth Above	1	2	3	2	3	4

Estimate is for the intercept term for receiving a higher rating from a linear regression with the provided bandwidths. Slopes are permitted to vary on either side of the cutoff. Sample is limited to schools that received an L, A, R, or E rating. Marginality below the rating cutoff is defined as affecting only one subject-group. Schools that receive ratings on appeal, were paired with another school, were identified as having undergone "special analysis" are excluded, as are schools with fewer than 200 students (see text). Robust standard errors clustered by school in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table 6: Regression Discontinuity Results for Effect of Rating on Expenditures and Staffing in TAAS

Outcome in Year t + 1 →	Instruction	Admin & Training	Counseling	Extra-Curricular	Regular Program	Gifted & Talented
	(1)	(2)	(3)	(4)	(5)	(6)
Low / Acceptable	346.6	7.6	-14.9	86.4***	197.6	-35.1
	(215.8)	(47.1)	(25.4)	(25.9)	(148.2)	(46.8)
Obs	336	1481	765	1093	525	8402
Bandwidth Below	3	4	3	4	4	2
Bandwidth Above	2	6	4	5	3	17
Acceptable / Recognized	24.9	19.9**	10.1*	1.8	-1.8	17.9**
	(46.6)	(9.4)	(5.3)	(9.5)	(43.7)	(7.6)
Obs	7271	5956	7271	7173	7271	8220
Bandwidth Below	3	3	3	1	3	3
Bandwidth Above	4	3	4	5	4	5
Recognized/Exemplary	83.8	40.2*	17.1***	-5.2	62.5	-23.6*
	(60.9)	(20.9)	(5.9)	(23.2)	(52.8)	(12.9)
Obs	3562	2443	5128	4232	5128	4232
Bandwidth Below	3	3	3	2	3	2
Bandwidth Above	2	1	4	3	4	3

Outcome in Year t + 1 →	Career & Tech (HS Only)	Special Intervention	Athletics (HS Only)	Student-Teacher Ratio	Student - Aide Ratio	Student-Admin Ratio
	(7)	(8)	(9)	(10)	(11)	(12)
Low / Acceptable	-501.9***	196.1	286.4**	-1.13	15.3	-34.0
	(156.7)	(179.3)	(122.8)	(0.91)	(18.8)	(35.3)
Obs	97	2584	102	736	515	331
Bandwidth Below	2	3	4	2	4	3
Bandwidth Above	3	8	3	4	3	2
Acceptable / Recognized	-11.1	-4.1	-32.5	-0.10	16.2	7.8
	(58.3)	(26.3)	(48.9)	(0.11)	(11.1)	(11.9)
Obs	853	5956	1145	4908	7112	4464
Bandwidth Below	3	3	3	1	3	3
Bandwidth Above	3	3	5	3	4	2
Recognized/Exemplary	-57.9	-26.1	-31.4	0.37*	-1.5	-14.0
	(70.3)	(30.1)	(44.5)	(0.22)	(6.8)	(15.0)
Obs	767	3562	889	3356	4293	4389
Bandwidth Below	3	3	3	2	3	3
Bandwidth Above	3	2	4	2	3	3

Estimate is for the intercept term for receiving a higher rating from a linear regression with the provided bandwidths. Slopes are permitted to vary on either side of the cutoff. Sample is limited to schools that received an L, A, R, or E rating. Marginality below the rating cutoff is defined as affecting only one subject-group. Schools that receive ratings on appeal, were paired with another school, were identified as having undergone "special analysiss," or have fewer than 200 students are excluded (see text). Robust standard errors clustered by school in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table 7: Regression Discontinuity Results for Effect of Rating on Expenditures and Staffing in TAKS

Outcome in Year t + 1 →	Instruction	Admin & Training	Counseling	Extra-Curricular	Regular Program	Gifted & Talented
	(1)	(2)	(3)	(4)	(5)	(6)
Low / Acceptable	-126.7	-54.6	29.8	-122.7*	42.8	-59.5
	(235.3)	(46.5)	(21.7)	(70.5)	(200.7)	(52.5)
Obs	2359	3013	1278	4667	2353	3802
Bandwidth Below	7	4	7	7	6	5
Bandwidth Above	4	5	2	7	4	6
Acceptable / Recognized	-83.0	0.6	20.5***	-7.2	-72.5	17.2*
	(52.8)	(9.4)	(7.6)	(19.9)	(44.8)	(8.9)
Obs	6021	6202	5768	5768	6202	6021
Bandwidth Below	4	5	3	3	5	4
Bandwidth Above	4	4	4	4	4	4
Recognized/Exemplary	32.8	4.0	9.1	20.6*	5.3	-24.0
	(140.2)	(26.7)	(13.5)	(10.9)	(113.2)	(31.2)
Obs	1463	1678	1678	1549	1334	1678
Bandwidth Below	3	3	3	2	2	3
Bandwidth Above	3	5	5	5	3	5

Outcome in Year t + 1 →	Career & Tech (HS Only)	Special Intervention	Athletics (HS Only)	Student-Teacher Ratio	Student - Aide Ratio	Student-Admin Ratio
	(7)	(8)	(9)	(10)	(11)	(12)
Low / Acceptable	-82.0	-96.4	-127.4	0.00	-12.4	-7.9
	(146.4)	(147.2)	(81.7)	(0.59)	(17.7)	(33.2)
Obs	614	1259	370	2358	2320	9091
Bandwidth Below	7	5	7	7	6	4
Bandwidth Above	4	2	2	4	4	13
Acceptable / Recognized	-16.5	4.7	-120.4*	0.41***	4.9	13.2
	(59.6)	(32.0)	(70.4)	(0.14)	(5.0)	(9.4)
Obs	488	6798	488	6018	7373	5975
Bandwidth Below	5	4	5	4	4	4
Bandwidth Above	3	5	3	4	6	4
Recognized/Exemplary	721.2	-77.2	506.2*	0.16	28.9***	7.7
	(439.0)	(81.6)	(285.3)	(0.30)	(9.0)	(19.1)
Obs	28	1177	27	1445	1534	1668
Bandwidth Below	3	2	3	2	3	3
Bandwidth Above	5	2	4	4	4	5

Estimate is for the intercept term for receiving a higher rating from a linear regression with the provided bandwidths. Slopes are permitted to vary on either side of the cutoff. Sample is limited to schools that received an L, A, R, or E rating. Marginality below the rating cutoff is defined as affecting only one subject-group. Schools that receive ratings on appeal, were paired with another school, were identified as having undergone "special analysis," or have fewer than 200 students are excluded (see text). Robust standard errors clustered by school in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table 8: Tests of Discontinuities in Exogenous Variables

Outcome in Year t →	Enrollment	Δ Enrollment from t to t+1	% Disadv	% Black	% Hispanic	% White
A. TAAS						
Low / Acceptable	75.7	-16	-5.1	2.2	-6.1	3.7
	(98.7)	(16.10)	(5.5)	(6.4)	(7.7)	(6.7)
Obs	531	381	531	531	531	531
Acceptable / Recognized	13.5	-4.2	0.1	1.9**	-0.1	-2.2
	(25.5)	(5.30)	(1.5)	(1.0)	(1.7)	(1.7)
Obs	8252	6295	8252	8252	8252	8252
Recognized/Exemplary	73.7***	-5.2	-1.5	0.9	-0.4	-0.3
	(23.6)	(5.50)	(1.7)	(1.0)	(1.8)	(2.0)
Obs	3579	2570	3579	3579	3579	3579
B. TAKS						
Low / Acceptable	-86.7	-13.3	2.5	1.0	4.6	-5.3
	(156.8)	(28.50)	(4.6)	(5.2)	(6.8)	(6.1)
Obs	2381	1578	2381	2381	2381	2381
Acceptable / Recognized	95.7***	-1.7	-0.2	-1.9*	3.7**	-2.3
	(22.6)	(4.90)	(1.6)	(1.0)	(1.8)	(1.9)
Obs	6048	4321	6048	6048	6048	6048
Recognized/Exemplary	5.3	14.2	-0.6	-1.4	1.9	-0.8
	(27.5)	(10.70)	(2.9)	(1.4)	(2.5)	(2.8)
Obs	1475	945	1475	1475	1475	1475

Estimate is for the intercept term for receiving a higher rating from a linear regression with the provided bandwidths. Slopes are permitted to vary on either side of the cutoff. Sample is limited to schools that received an L, A, R, or E rating. Marginality below the rating cutoff is defined as affecting only one subject-group. Schools that receive ratings on appeal, were paired with another school, were identified as having undergone "special analysis," or have fewer than 200 students are excluded (see text). We use the leave-one-out cross-validation optimal bandwidths for per-student total expenditures. See table 4 for the bandwidths. Results are similar using the median optimal bandwidth across all outcomes. Robust standard errors clustered by school in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table 9: "Ratings Shock" Estimates of Resouces on Ratings Changes in Transition Year

Δ from 2002-03 to 2003-04 →	Total	Instruction	Admin & Training	Counseling	Extra-Curricular	Regular Program	Gifted & Talented
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Estimated Rating Falls to L	-14.9	6.9	-1.0	-2.1	4.2	38.6	-0.7
	(34.8)	(27.6)	(4.7)	(3.0)	(3.1)	(28.3)	(3.2)
Estimated Rating Falls to A	45.8*	45.3**	3.5	0.4	-3.3	60.0***	-3.0
	(25.6)	(20.1)	(4.2)	(2.4)	(2.3)	(21.4)	(3.5)
Estimated Rating Falls to R	-2.6	10.5	0.3	1.9	-3.5	16.7	-7.0*
	(32.4)	(25.2)	(4.9)	(3.0)	(2.7)	(28.3)	(4.1)
Estimated Rating Increases	-15.2	-12.7	-1.4	0.0	2.2	-24.5	-8.8
	(64.9)	(56.9)	(8.9)	(5.4)	(3.3)	(57.4)	(7.1)
Low Rating in 2001-02	97.5	11.2	52.1***	18.9**	4.1	-40.7	0.7
	(126.1)	(114.4)	(17.5)	(7.8)	(5.1)	(123.0)	(8.7)
Acceptable Rating in 2001-02	93.4***	83.5***	12.3**	5.0*	-2.9	88.2***	-5.5
	(31.7)	(25.1)	(5.0)	(3.0)	(2.9)	(30.0)	(3.9)
Recognized Rating in 2001-02	24.4	12.9	8.2***	3.4*	1.4	21.8	-2.9
	(19.9)	(15.0)	(3.0)	(1.8)	(2.2)	(20.5)	(2.6)
Observations	5142	5142	5142	5142	5142	5142	5142

Δ from 2002-03 to 2003-04 →	Career & Tech (HS Only)	Special Interest	Athletics (HS Only)	Student-Teacher Ratio	Student - Aide Ratio	Student-Admin Ratio
	(8)	(9)	(10)	(11)	(12)	(13)
Estimated Rating Falls to L	22.3	-3.8	21.9	-0.14*	0.5	4.2
	(22.1)	(17.2)	(15.7)	(0.07)	(3.8)	(14.9)
Estimated Rating Falls to A	38.4	-28.0*	20.2	-0.26***	-3.4	16.9
	(30.2)	(16.5)	(18.7)	(0.06)	(4.6)	(11.1)
Estimated Rating Falls to R	6.1	-22.2	3.8	-0.22	6.9	-0.6
	(41.2)	(21.2)	(22.4)	(0.15)	(6.3)	(10.8)
Estimated Rating Increases	-150.5	-25.4	-25.4	0.14	-3.7	8.4
	(141.5)	(24.6)	(24.8)	(0.13)	(5.5)	(10.1)
Low Rating in 2001-02	212.6	50.5	54.7*	-0.41*	6.3	-1.6
	(143.7)	(59.1)	(28.5)	(0.22)	(10.5)	(20.7)
Acceptable Rating in 2001-02	32.4	-49.7**	22.3	-0.26***	-2.3	6.1
	(32.7)	(24.8)	(17.7)	(0.08)	(5.2)	(13.3)
Recognized Rating in 2001-02	3.6	-25.0*	5.1	-0.02	6.8	-11.5
	(15.6)	(14.4)	(7.6)	(0.04)	(6.2)	(8.3)
Observations	845	5142	845	5141	4961	5072

Sample is limited to schools that received an L, A, R, or E rating in both 2001-02 & 2003-04. Schools that received ratings on appeal, were paired with another school, were identified as having undergone "special analysis," or have fewer than 200 students are excluded (see text). Regressions also include controls for % of students in each grade level, % black, % Hispanic, % Asian, % Native American, & % economically disadvantaged. Robust standard errors clustered in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table 10: School Fixed Effects Regressions of Resouces on Ratings - TAAS

Outcome in Year t + 1 →	Total	Instruction	Admin & Training	Counseling	Extra-Curricular	Regular Program	Gifted & Talented
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Low	-3.4	-6.4	9.6	7.2*	-4.6*	51.0	-2.4
	(57.5)	(48.4)	(10.3)	(4.0)	(2.6)	(36.2)	(5.4)
Recognized	7.0	5.1	2.9	1.9**	1.6	1.7	-0.2
	(10.2)	(7.5)	(1.9)	(0.9)	(1.0)	(8.2)	(1.6)
Exemplary	9.5	-4.5	9.1	3.7***	4.4***	-0.5	2.0
	(17.1)	(11.8)	(5.8)	(1.4)	(1.7)	(12.8)	(2.4)
Observations	26,560	26,560	26,560	26,560	26,560	26,560	26,560

Outcome in Year t + 1 →	Career & Tech (HS Only)	Special Intervention	Athletics (HS Only)	Student-Teacher Ratio	Student -Aide Ratio	Student-Admin Ratio
	(8)	(9)	(10)	(11)	(12)	(13)
Low	2.8	-7.2	-14.2*	-0.082	-7.3*	-9.13
	(19.2)	(23.5)	(8.4)	(0.115)	(4.3)	(8.99)
Recognized	-9.4	1.1	-0.9	0.010	-3.2	3.69
	(6.8)	(4.6)	(5.3)	(0.026)	(2.4)	(3.92)
Exemplary	-7.6	-4.0	-4.3	0.003	0.0	0.26
	(9.8)	(6.5)	(11.4)	(0.036)	(3.2)	(4.95)
Observations	4,248	26,560	4,248	26,541	25,932	26,353

Sample is limited to schools that received an L, A, R, or E rating. Schools that received ratings on appeal, were paired with another school, were identified as having undergone "special analysis," or have fewer than 200 students are excluded (see text). Regressions also include controls for % of students in each grade level, % black, % Hispanic, % Asian, % Native American, % economically disadvantaged, and year dummies. Robust standard errors clustered by school in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table 11: School Fixed Effects Regressions of Resouces on Ratings - TAKS

Outcome in Year t + 1 →	Total	Instruction	Admin & Training	Counseling	Extra-Curricular	Regular Program	Gifted & Talented
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Low	91.0**	51.1	22.7***	6.9**	0.9	44.4	0.1
	(42.9)	(31.2)	(7.4)	(2.8)	(3.3)	(36.9)	(3.0)
Recognized	4.1	0.2	1.5	0.8	-2.8**	6.3	2.1
	(11.8)	(8.8)	(1.9)	(1.1)	(1.1)	(12.1)	(1.4)
Exemplary	41.0	29.3	-0.6	4.9*	-3.7*	-0.3	6.5*
	(26.0)	(19.0)	(4.3)	(2.7)	(1.9)	(20.2)	(3.4)
Observations	23,168	23,168	23,168	23,168	23,168	23,168	23,168

Outcome in Year t + 1 →	Career & Tech (HS Only)	Special Intervention	Athletics (HS Only)	Student-Teacher Ratio	Student - Aide Ratio	Student-Admin Ratio
	(8)	(9)	(10)	(11)	(12)	(13)
Low	-4.3	71.2***	-14.7	-0.204***	4.8	-23.43**
	(15.6)	(15.8)	(11.6)	(0.073)	(4.1)	(9.77)
Recognized	4.0	1.1	3.1	0.014	4.3	-2.47
	(11.2)	(5.9)	(12.7)	(0.038)	(3.4)	(6.73)
Exemplary	-51.2	13.6	25.4	-0.025	5.0	0.10
	(34.1)	(10.6)	(41.0)	(0.054)	(4.8)	(7.83)
Observations	3,679	23,168	3,679	23,153	22,728	23,038

Sample is limited to schools that received an L, A, R, or E rating. Schools that received ratings on appeal, were paired with another school, or were identified as having undergone "special analysis," have fewer than 200 students are excluded (see text). Regressions also include controls for % of students in each grade level, % black, % Hispanic, % Asian, % Native American, % economically disadvantaged, and year dummies. Robust standard errors clustered by school in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table 12 - Regressions of Resources on Whether School Is Marginal For a Rating - TAAS

School Sample	Outcome in Year t + 1 → Estimate	Total (1)	Instruction (2)	Admin & Training (3)	Counseling (4)	Extra-Curricular (5)	Regular Program (6)	Gifted & Talented (7)
Acceptable	Marginal L/A Boundary	68.9*	39.1	14.0*	-0.5	3.5	62.5**	-1.4
		(39.4)	(29.9)	(8.0)	(3.6)	(2.6)	(31.7)	(3.7)
	Marginal A/R Boundary	13.6	8.9	0.5	-0.9	2.5	18.5	-2.4
		(20.0)	(14.5)	(3.2)	(1.7)	(1.9)	(15.2)	(2.2)
	Observations	12111	12111	12111	12111	12111	12111	12111
Recognized	Marginal A/R Boundary	11.5	15.6	-5.0	0.7	0.2	-2.7	1.6
		(26.4)	(17.1)	(12.1)	(1.9)	(1.8)	(17.9)	(2.6)
	Marginal L/A Boundary	37.6	33.7*	-5.0	4.22*	2.1	34.43*	4.0
		(31.0)	(19.4)	(15.9)	(2.3)	(2.6)	(20.4)	(3.5)
	Observations	8614	8614	8614	8614	8614	8614	8614

School Sample	Outcome in Year t + 1 → Estimate	Career & Tech (HS Only) (8)	Special Intervention (9)	Athletics (HS Only) (10)	Student-Teacher Ratio (11)	Student - Aide Ratio (12)	Student-Admin Ratio (13)
Acceptable	Marginal L/A Boundary	73.4*	3.1	3.0	-0.05	-2.5	-0.2
		(37.6)	(21.0)	(9.4)	(0.11)	(4.5)	(10.1)
	Marginal A/R Boundary	-12.7	8.6	-1.8	-0.02	-1.4	3.8
		(9.9)	(9.2)	(9.1)	(0.08)	(7.4)	(4.9)
	Observations	1961	12111	1961	12099	11860	12031
Recognized	Marginal A/R Boundary	3.8	17.1	-5.2	0.00	2.6	-4.7
		(15.9)	(16.1)	(16.0)	(0.05)	(4.9)	(5.4)
	Marginal L/A Boundary	-10.8	-1.3	-26.9	-0.08	1.1	-9.2
		(22.5)	(11.5)	(26.4)	(0.06)	(3.6)	(9.4)
	Observations	1231	8614	1231	8612	8433	8548

A school is considered marginal if it is within 3 students-group-subject units of a boundary. Results using 5 students were simliar and are provided in the online appendix. Sample is limited to schools that received an L, A, R, or E rating. Schools that received ratings on appeal, were paired with another school, were identified as having undergone "special analysis," or have fewer than 200 students are excluded (see text). Regressions also include controls for % of students in each grade level, % black, % Hispanic, % Asian, % Native American, % economically disadvantaged, year dummies, and school fixed-effects. Robust standard errors clustered by school in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table 13 - Regressions of Resources on Whether School Is Marginal For a Rating - TAKS

School Sample	Outcome in Year t + 1 → Estimate	Total (1)	Instruction (2)	Admin & Training (3)	Counseling (4)	Extra-Curricular (5)	Regular Program (6)	Gifted & Talented (7)
Acceptable	Marginal L/A Boundary	48.8**	25.8	7.43*	0.7	2.8	-0.2	0.3
		(22.97)	(17.23)	(4.03)	(1.99)	(2.69)	(16.75)	(2.18)
	Marginal A/R Boundary	5.0	-10.2	3.8	4.53*	-4.63*	-2.3	0.0
		(27.0)	(19.6)	(4.4)	(2.6)	(2.7)	(19.9)	(2.2)
	Observations	12537	12537	12537	12537	12537	12537	12537
Recognized	Marginal A/R Boundary	-24.1	-16.9	-3.6	-3.5*	0.3	-23.1	0.2
		(20.4)	(15.1)	(3.8)	(2.0)	(1.5)	(16.7)	(2.5)
	Marginal L/A Boundary	27.7	22.4	0.9	-1.1	-1.0	10.4	-3.8
		(34.1)	(26.0)	(5.1)	(3.7)	(2.5)	(35.8)	(5.1)
	Observations	7985	7985	7985	7985	7985	7985	7985

School Sample	Outcome in Year t + 1 → Estimate	Career & Tech (HS Only) (8)	Special Intervention (9)	Athletics (HS Only) (10)	Student-Teacher Ratio (11)	Student - Aide Ratio (12)	Student-Admin Ratio (13)
Acceptable	Marginal L/A Boundary	-3.0	23.3**	10.7	-0.05	11.0	11.6
		(12.2)	(9.2)	(16.1)	(0.05)	(12.8)	(18.7)
	Marginal A/R Boundary	14.6	3.6	-24.06*	0.00	1.4	-4.2
		(15.3)	(11.8)	(13.4)	(0.05)	(3.2)	(8.3)
	Observations	2183	12537	2183	12527	12341	12474
Recognized	Marginal A/R Boundary	-7.0	-0.2	6.3	0.16	-8.2	-0.5
		(27.1)	(10.4)	(37.3)	(0.23)	(9.5)	(5.1)
	Marginal L/A Boundary	17.9	6.9	-31.1	-0.04	-1.4	4.2
		(51.4)	(27.7)	(35.0)	(0.08)	(2.7)	(7.2)
	Observations	405	7985	405	7982	7803	7934

A school is considered marginal if it is within 3 students-group-subject units of a boundary. Results using 5 students were simliar and are provided in the online appendix. Sample is limited to schools that received an L, A, R, or E rating. Schools that received ratings on appeal, were paired with another school, were identified as having undergone "special analysis," or have fewer than 200 students are excluded (see text). Regressions also include controls for % of students in each grade level, % black, % Hispanic, % Asian, % Native American, % economically disadvantaged, and year dummies. Robust standard errors clustered by school in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Appendix Table 1: Requirements for TAAS Accountabiltiy Ratings

Subject	Math, Reading	Writing	Social Studies	Drop-Outs	Attendance
Grades	3 - 8, 10	4, 8, 10	8	7 - 12	All
Groups	White, Black, Hisp, Econ Dis, All	White, Black, Hisp, Econ Dis, All	All only	White, Black, Hisp, Econ Dis, All	White, Black, Hisp, Econ Dis, All
			A. Acceptable		
1998-99	40% or RI	40% or RI	n/a	6% or RI	94%
1999-00	45%	45%	n/a	6%	94%
2000-01	50%	50%	n/a	5.50%	94%
2001-02	50%	50%	n/a	5%	-
2002-03	55%	55%	50%	5%	-
			B. Recognized		
1998-99	80%	80%	n/a	3.50%	94%
1999-00	80%	80%; 3-8, 10	n/a	3.50%	94%
2000-01	80%	80%; 3-8, 10	n/a	3.00%	94%
2001-02	80%	80%; 3-8, 10	n/a	2.50%	-
2002-03	80%	80%; 3-8, 10	80%	2.50%	-
			C. Exemplary		
1998-99	90%	90%	n/a	1%	94%
1999-00	90%	90%; 3-8, 10	n/a	1%	94%
2000-01	90%	90%; 3-8, 10	n/a	1%	94%
2001-02	90%	90%; 3-8, 10	n/a	1%	-
2002-03	90%	90%; 3-8, 10	90%	1%	-

RI - Required improvement. Schools that do not meet the requirement could get the higher rating by showing sufficient increase in the performance measure.

Notes: To count, all of the subject/student group combinations must be at least either: 30 students & 10% of the student body, or 200 students (prior to 2001) / 50 students (2001 and later).

Appendix Table 2: Requirements for TAKS Accountabiltiy Ratings

Subject	Math	Reading/ELA	Writing	Social Studies	Science	SDAA	Drop-Outs	Completions
Grades	3 - 11	3 - 11	4, 8, 10	8, 10, 11	5, 10, 11	3 - 11	7 - 8	12
Groups	White, Black, Hisp, Econ Dis, All	White, Black, Hisp, Econ Dis, All	White, Black, Hisp, Econ Dis, All	White, Black, Hisp, Econ Dis, All	White, Black, Hisp, Econ Dis, All	All	White, Black, Hisp, Econ Dis, All	White, Black, Hisp, Econ Dis, All
				A. Acceptable				
2003-04	35% or RI	50% or RI	50% or RI	50% or RI	25% or RI	50% or RI	1% or RI	75% or RI
2004-05†	35% or RI	50% or RI	50% or RI	50% or RI	25% or RI	50% (no RI - new exam)	1% or RI	75% or RI
2005-06	40% or RI	60% or RI	60% or RI	60% or RI	35% or RI	50% or RI	1% (no RI - new calc method)	75% or RI
2006-07	45% or RI	65% or RI	65% or RI	65% or RI	40% or RI	50% or RI	2% or RI	75% or RI
				B. Recognized				
2003-04	70% or RI	70% or RI	70% or RI	70% or RI	70% or RI	70% or RI	0.7% or RI	85%
2004-05	70% or RI	70% or RI	70% or RI	70% or RI	70% or RI	70% (no RI - new exam)	0.7% or RI	85% or RI
2005-06	70% or RI	70% or RI	70% or RI	70% or RI	70% or RI	70% or RI	0.7% (no RI - new calc method)	85% or RI
2006-07	75% or RI	75% or RI	75% or RI	75% or RI	75% or RI	70% or RI	0.7% or RI	85% or RI
				C. Exemplary				
2003-04	90%	90%	90%	90%	90%	90%	0.2%	95%
2004-05	90%	90%	90%	90%	90%	90%	0.2%	95%
2005-06	90%	90%	90%	90%	90%	90%	0.2%	95%
2006-07	90%	90%	90%	90%	90%	90%	0.2%	95%

† Although the passing rates did not increase in 2004-05 the standard for whether individual students passed did increase.

RI - Required improvement. Schools that do not meet the requirement could get the higher rating by showing sufficient increase in the performance measure.
SDAA - State Developed Alternative Assessment - Test for certain special education students. Passing rates based on percent of tests taken.
ELA - English Language Arts

Notes: To count, all of the subject/student group combinations must be at least either: 30 students, or 10% of the student body, or 50 students. Schools are also granted exceptions for for a certain number of low-scoring subject/group combinations based on the total number of subject/groups that count towards the rating. Exceptions can only increase a rating from L to A.