The Role of School Accountability and Charter Schools in the Achievement of Public School Students

BY

PATRICK LAWRENCE BAUDE
B.S.E., University of Pennsylvania, 2003
M.S., State University of New York at Buffalo, 2010

THESIS

Submitted as partial fulfillment of the
Requirements for the degree of Doctor of Philosophy in Economics
in the Graduate College of the
University of Illinois at Chicago, 2017

Chicago, Illinois

Defense Committee:

Steven Rivkin, Chair and Advisor
Joseph Persky
Marcus Casey
Ben Ost
Mark Smylie, Professor Emeritus, Department of Educational Policy Studies

This thesis is dedicated to Petra, Vera and Emil, for providing motivation and teaching me lessons in persistence and perspective during the completion of my graduate studies.

ACKNOWLEDGEMENTS

I would like to thank my thesis committee for their support and assistance. In particular, I owe particular gratitude to Steve Rivkin and Marcus Casey. Their advice, mentorship, and friendship have been invaluable.

In addition, I would like to acknowledge Greg Branch, Rodney Andrews, Mark Lu and Greg Phelan at the Texas Schools Project at the University of Texas at Dallas where portions of the data used in this thesis are housed. Their help with navigating the data holdings during my visits to the data center in Dallas were instrumental in the completion of my work. This work was done in conjunction with the Texas Schools Project at the University of Texas at Dallas. Ross Cole provided superb research assistance. The conclusions of this research do not necessarily reflect the opinions or official position of the Texas Education Agency, the Texas Higher Education Coordinating Board, or the State of Texas.

Further, I would like to thank three anonymous referees for valuable feedback, and seminar participants at the following conferences: 2014 Association for Public Policy Analysis and Management Fall Research Conference; 2014 Association for Education, Finance, and Policy Annual Conference; and the 2014SKILLS workshop at the Center for Economic Research and Graduate Education and the Economics Institute in Prague, CZ.

PLB

CONTRIBUTION OF AUTHORS

The components of this dissertation relating to the evolution of charter school quality come directly from joint work with Marcus Casey, Eric Hanushek and Steven Rivkin. An earlier draft of this paper was entitled "The Evolution of Charter School Quality" and was circulated as National Bureau of Economic (NBER) Working Paper 20645. The NBER does not hold the copyright for its working papers. This project has been presented at several economics conferences, including by myself at the Association for Public Policy Analysis and Management Fall Research conference in Albuquerque, NM on November 6, 2014.

All authors contributed to the evolution of the conception and design of the study. I conducted the data analysis at the Texas Schools Project located at The University of Texas at Dallas. Drafting of the article was primarily done by Steve Rivkin and Marcus Casey with Rick Hanushek and myself conducting critical revision to the drafts. All authors (Patrick Baude, Marcus Casey, Rick Hanushek, and Steve Rivkin) have approved the final version of this paper and acknowledge its inclusion in this dissertation.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF TABLES

LIST OF TABLES (Continued)

LIST OF ABBREVIATIONS

AEA	Alternative Education Accountability
AEC	Alternative Education Campus
CMO	Charter management organization
ELA	English language arts
FE	Fixed Effects
LATE	Local average treatment effect
LLR	Local linear regression
MSR	Minimum size requirement
NAEP	National Assessment of Educational Progress
NBER	National Bureau of Economic Research
NCES	National Center for Education Statistics
OLS	Ordinary least squares
PEG	Public education grant
RD	Regression discontinuity
RI	Required improvement
SDAA	Stata Developed Alternative Assessment
TAAS	Texas Assessment of Academic Skills
TAKS	Texas Assessment of Knowledge and Skills
TEA	Texas Education Agency
TPM	Texas projection measure
TPS	Traditional Public School
VA	Value Added

SUMMARY

This dissertation investigates two education policy reforms which strengthen incentives and choice into the public education system in the state of Texas. These policies are school accountability systems and charter schooling. School accountability systems evaluate schools and publish ratings so as to make information about the standards-based performance more readily accessible to the public and state and local administrators. Charter schools are publicly funded schools operated by private non-profit organizations. Charter schools are held to the same accountability standards as traditional public schools while not being subject to the same rules and regulations. In this way charter schools can offer families a choice between pedagogical approaches when deciding where to send their children to school. By increasing the availability of information about school quality and expanding the choices available to families, these policies are intended to improve educational outcomes through the market forces.

Empirical results show that the receipt of a lower accountability rating is found to cause schools to improve achievement the following year only when that lower rating is coupled with state sanctions. When a rating difference incurs no statutory intervention, there is no causal effect on subsequent academic achievement. Despite the significant achievement response at boundaries that trigger sanctions, no causal effect is found on the tendency of families to exercise school choice by moving their children to a new school. Combined, these results suggest that ratings provide little additional information to families about school quality and that schools are not incentivized to increase their rating unless it risks statutory punishment.

On the charter front, the quality of schools operated under charters, as measured by valued-added to test scores, increased relative to traditional public schools over the ten-year period from 2001-2011, This increase in average quality came with a decrease in the variance in quality. The closure of low performing schools, the improvement of existing schools, and expansion of higher-performing charter management organizations increased charter effectiveness relative to traditional public schools. There is evidence that reduced student turnover and an increased share of charters adhering to No Excuses-style curricula contribute to these improvements.

I. INTRODUCTION

This dissertation evaluates two related education policy reforms using administrative data from the state of Texas. These policies, school accountability and charter schooling, are similar in that they do not prescribe changes to pedagogy nor curriculum. Rather, these policies help create an environment conducive to innovation where excellence is rewarded and underperformance can be identified. The increased competition and reduced information frictions that arise from these policies make it appropriate to evaluate them from an economic perspective.

Texas school accountability legislation was passed in 1993 and, by leveraging pre-existing student data management architecture and curriculum-based standardized testing, created a mandate that schools be assigned ratings based on student academic performance along with other factors. The state education agency had several goals in the design and implementation of the accountability rating system. Most notably, the system is intended to aid in the identification of schools in need of reform or recognition, and to provide information to communities and educators about quality differences between schools. During the period of analysis in this dissertation, while schools were assigned one of four ordered ratings, statutory consequences were only triggered when a school was assigned the lowest rating. These statutory effects, combined with any potential reputational effects at all rating levels, are intended to provide an incentive for schools to increase achievement and therefor their rating. In this way, a stable and transparent system for assigning accountability ratings provides an objective mechanism to ensure a minimum standard of quality across all public schools in the state through either formal intervention (sanctions) or market forces (reputational effects). Further, the availability of reliable and objective information about school quality is particularly important when low cost school choice options are available. While a school's reputation is important to administrators and local politicians for reasons unrelated to enrollment, the reputational effect of receiving a low rating does not directly affect the demand for a school unless families have the ability to exercise school choice.

Conventional school choice is exercised via the residential choice decision. Traditional public school enrollment options are based on a family's location of residence. As anyone who has purchased a home knows, the residential location and school choice decisions are very much a joint decision. In this way conventional school choice results from Tiebout sorting by inter-district moves. Changing schools via this mechanism, however, can be costly both monetarily and in terms of time.

In Texas, another major source of school choice comes in the form of charter schools. Charter school legislation in Texas was introduced in 1995 with the first schools beginning operation in 1996. Charter schools are publicly funded schools that are privately operated and face relaxed operational constraints compared to traditional public schools. These schools do not draw students from traditional catchment areas, but rather they operate within the catchment areas of existing traditional public schools. Enrollment is determined through an open application process with a lottery should there be more applications than available seats. By 2011 almost 5% of public school students in Texas attended one of the more than 475 state authorized charter school campuses.

With fewer operational constraints, charter school administrators are able to implement alternative approaches to educating students that are not possible at traditional public schools. In this way, charter school laws create a deregulated subsector within the public education system which is more conducive to innovation. Some areas where charters have differentiated themselves from traditional public schools are in hiring practices, attitudes towards discipline, varying the length of the school day or year, and the implementation of specialized curricula.

It is important to note that the differences between charters and traditional public schools are not limited to campus-level operations. Many charter school campuses are part of larger entities called charter management organizations (CMOs) which can operate dozens of individual sites. CMOs are able to take advantage of economies of scale along various dimensions of school operations and often implement both horizontal and vertical integration between school sites. These economies are not limited to curriculum development but also include back office operations like reporting and accreditation, professional development, and the establishment of school feeder networks. Since decision making and outcomes can occur at both the individual campus level and the broader CMO level, this dissertation presents analyses at both levels.

Importantly charter students still sit for the annual statewide standardized tests, and charter schools are still assigned ratings under the state school accountability rating system based primarily on the performance of students on these tests. Ultimately, while charters and accountability are separate reforms, they both aim to improve education by introducing and strengthening market pressures within the public education system through increased competition and reduced frictions in the dissemination of information about school quality.

A. School Accountability

The Texas Education Agency lists several guiding principles behind the construction of their accountability rating system. These include the improvement of student performance, 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. I evaluate the school accountability policy in terms of these stated guiding principles. Before outlining a framework for thinking about how ratings achieve these goals, I will lay out how my contributions fit within the existing literature on the effects of school accountability ratings.

The existing literature falls into three main areas. The first measures the effects of the imposition of an accountability policy in general. These papers use a before and after or differences in differences type strategy. Another strand looks at the effect of changing the salience of an assigned rating on school choice decisions. These papers make use of variation in the timing of the release of ratings or involve field experiments. The final strand estimates the effect of being assigned a certain rating instead of another.

Prior findings in this final strand focus on the effects on achievement and are mixed. Theory predicts that reputational effects and the threat of interventions give administrators incentive to improve when a low rating is assigned. Positive test score effects of a shock to the lowest rating level have been found in Florida, New York City, and North Carolina, while other work has found that a failing rating in

California results in a decline in future achievement. This paper contributes to this body of evidence by estimating rating effects in a new policy environment, Texas.

Additionally, this thesis combines the second and third strands of the literature by estimating effects of the assigned rating on the family decision about where to send children to school, hereafter the school choice decision. Ratings provide information to families about public school quality. As pointed out by Loeb et al (2011), the availability and reliability of information about school quality are fundamental factors affecting the demand schools. I compare the decision of families between schools that receive different information signals about school quality. 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.

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) find that while the charter school reenrollment decision in Texas has been shown to be positively related to school quality as measured by either 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.

Advocates of school accountability systems hope that the market forces introduced by assigning ratings to schools will increase the quality of public education. However, since families and administrators already have information about the quality of schools from other sources, it is difficult to know if new information is learned from a rating and if so, how this information affects behaviors. In this paper, I estimate school fixed effects models and a regression discontinuity (RD) design to identify two effects of the information about school quality contained in ratings. The first is the degree to which rating assignment leads to improved performance. I add to the existing evidence on this question by estimating the effect in the state of Texas, which operates one of the largest public school systems in the country. The literature differs across policy environments, but the bulk of the evidence, my results included, indicates that ratings have a positive effect on test scores at the bottom of the school quality distribution and little effect at other points in the distribution. The second is how sensitive the family school choice decision is to both the school quality signal embodied by a rating and to the effects of any punitive interventions associated with a rating. While other work has estimated the effects on school choice of the existence or salience of a rating system as a whole, this is the first paper to my knowledge that estimates the effects on school choice of the specific rating assigned to a school.

In Texas during the years analyzed, schools were assigned one of four ratings (in decreasing order): *exceptional*, *recognized*, *acceptable*, or *unacceptable*. Higher ratings are associated with smaller schools, higher average test scores, and higher proportions of minority and low income students. Large mean differences in these observable school characteristics across rating levels suggest equally large differences in unobservable characteristics. In order to identify the causal effects of receiving a certain rating it is necessary to use an empirical strategy that addresses these unobservable differences. I first estimate school fixed effects (FE) models which control for time-invariant unobservable characteristics that differ between schools. The identifying variation in these models comes from schools that experience a rating change at some point during the period of analysis. School quality and test scores, and therefor ratings, are determined by both school and family characteristics. After controlling for time-varying demographic characteristics and school achievement in the FE analysis, any response to changes in ratings are likely a combination of changes to school effectiveness, transitory shocks and measurement error. To address these concerns, I then estimate the rating effect using a fuzzy regression discontinuity (RD) design.

The fuzzy RD design compares schools on opposite sides of the boundaries between ratings where the assigned rating is plausibly random. This design isolates rating differences that represent changes to the likelihood of intervention and to the information signal about school quality that are orthogonal to all other school characteristics, including unobservables. This rating variation at the margin between rating levels allows produces more compelling estimates of the effects of ratings on school choice and academic performance. I construct two different running variables that introduce discontinuities in the relationship between school characteristics and the assigned ratings. The first discontinuity is based on the strict achievement cutoffs for each rating level, while the second results from the greater number of criteria that are imposed on larger and more diverse schools. These discontinuities cause schools that are marginally different to experience substantially different probabilities of being assigned different ratings. Neither theory, nor empirical evidence, suggest that schools have the ability to precisely manipulate their position relative to the discontinuity boundaries thereby allowing these estimates to take on a casual interpretation.

FE estimates indicate that school-level math pass-rates are over 4.5 percentage points higher the next year when a school receives the lowest rating compared to when receiving the highest rating, with no difference for reading. The math pass-rate difference is consistent with school administrators responding to incentives to improve their reputations and to sanctions triggered by consecutive low ratings, while the lack of effects in reading is not uncommon in the literature as math scores tend to be more responsive in the short run to policy differences. I then use the school FE model to estimate the effects of ratings on reenrollment rates. Two-year ahead reenrollment is 2.8 percentage points higher when a school is rated *exemplary* compared to when it is rated *unacceptable*. These differences are suggestive that the negative information signal associated with a low rating outweighs any expected achievement gains resulting from interventions. At charter schools, test scores tend to be more responsive though less precisely estimated, and the reenrollment differences that take two years to appear at traditional schools occur in the first year at charters. Both of these differences are consistent with stronger incentives in charter sector due to more competition facing and more mobile families.

In the RD analysis, I find that achievement is most sensitive to the assigned rating at the *acceptable-unacceptable* boundary with *unacceptable* schools experiencing a significant 7-9 percentage point increase in their math pass rate the following year. As in the FE analysis, RD results for reading are insignificant at all boundaries and again are generally smaller than the estimates for math. While noisy, the general pattern is that lower ratings trigger marginally lower reenrollment rates. This is consistent with families either not learning much new information from or not being very responsive to serendipitous rating differences. Compared to the effects of class size reductions, It is important to note that RD estimates are based on difference between schools at the margin between adjacent rating levels. While internally valid, conclusions drawn from the RD estimates may not be applicable to schools away from the margins. I mitigate this drawback of the RD designs by estimating RD models at six different points in the school quality distribution. Importantly, the differences between estimates at the different margins are consistent with the differences in the incentives that exist at the different RD boundaries. These different LATEs allow for the RD approach to produce more generalizable estimates of the effects of rating assignment.

On the other hand, the school FE estimates fail to fully account for time varying unobservable school characteristics correlated with ratings, and are potentially plagued by mean reversion. Mean reversion may occur for several reasons. Some of these reasons include classical measurement error in test scores, real external shocks to schools from one year to the next such as extreme weather or widespread illness on or around the testing day, and real fluctuations in teacher or administrator productivity leading to school quality fluctuations. Importantly, permanent changes to school quality do not result in mean reversion. If present, mean reversion from any of these sources may contaminate the estimate of the effect of ratings on the future pass rate, but not the future reenrollment rate. This fact is due to the roll pass rates play in the determination of ratings, while reenrollment does not figure into the rating assignment mechanism. On the whole, the evidence from both approaches support the conclusion that ratings promote test score improvements and that families school choices are not significantly impacted by any information learned from ratings.

B. Charter School Quality

The rapid expansion of the charter school sector in many states has proved controversial, in part because of mixed evidence on the role charters play in improving academic achievement. Some studies focusing on oversubscribed urban charter schools have found positive achievement impacts,[1] though these findings have yet to be generalized to other settings. In particular, studies focusing on all charter schools in a geographic area, not just those that are over-subscribed, have found much smaller or even negative impacts of charter schools on achievement.[2] These incongruous findings support the claims of both advocates and opponents of charter schools. Advocates point to the high quality of many oversubscribed schools as evidence that charters are living up to their promise. Opponents highlight the mediocre average outcomes and high variability in performance among the broader set of schools as evidence against further charter sector expansion or at times even the continuation of charter schools. Moreover, drawing appropriate implications for policy from existing evidence is hampered by the cross-sectional nature of the analyses. Deeper understanding of this market-oriented reform requires examination of the longer-term dynamics of the charter sector.

Although little comprehensive research exists on the evolution of charter school quality, two studies provide evidence consistent with effective market forces pushing schools to improve. First, Hanushek et al. (2007) show that higher school value-added increases the probability of student reenrollment in charter schools, suggesting that households respond to quality. Second, CREDO (2013) finds that average charter school effectiveness has improved relative to traditional public schools in a number of states. It highlights the closure of poorly performing charter schools as an important mechanism for improvement.

This paper capitalizes on detailed longitudinal data for students and schools to contribute new evidence to this debate. It has two principal aims. First, the paper describes how the distribution of charter school quality in Texas, one of the largest charter school states, has evolved between 2001 and 2011. Second, it investigates the extent to which more fundamental factors –student mobility, student selection

---

[1] Abdulkadiroğlu et al. (2011), Angrist et al. (2012), and Angrist, Pathak, and Walters (2013) report results for charter schools in and around Boston, and Dobbie and Fryer (2011) and Hoxby, Murarka, and Kang (2009) report results for New York City.

[2] See, for example, evidence from statewide studies in Bifulco and Ladd (2006), Sass (2006), Booker et al. (2007), and Hanushek et al. (2007). See also the multiple state comparisons in CREDO (2009, (2013).

into and out of charters, and the share of schools that adhere to a "No Excuses" philosophy – contribute to the observed changes in school quality.

The descriptive analysis provides strong evidence that charter school quality has improved over time in Texas. Specifically, we find that mean charter school mathematics value-added improved relative to traditional public schools by approximately 0.12 of a standard deviation, and similar improvements were seen throughout the distribution. Improvement in mean charter school reading value-added was about 0.09 of a standard deviation; however, for reading this improvement was concentrated at the lower end of the charter school quality distribution. Declines in the variances of charter-school mathematics and reading value-added driven by dramatic reductions in the numbers of schools in the left tales of the respective distributions accompany improvements in the average performance of charter schools.

To understand better the source of these improvements, we first consider how the dynamics of school entry and exit affect the distribution of school quality. We find that the voluntary and involuntary closure of underperforming schools, the increase in the quality of new entrants, and the improvement of existing schools combine to increase the mean and to reduce the variance of charter school value-added relative to traditional public schools. First, similar to the findings in CREDO (2013), schools that close prior to 2011, either voluntarily or following state authorizer intervention, come disproportionately from the lower end of the quality distribution. Second, charter schools that open after 2001 and are still operating in 2011 have an average value-added that far exceeds those that closed and roughly equals the average of charter schools open the entire period. Third, average value-added increases for charter schools that remain open throughout the decade. Although average school value-added for charters and traditional public schools is quite similar in 2011, the changes over time suggest that market forces contribute to these dynamic improvements in the charter sector.

The analysis of school factors provides evidence that the increase in the share of charter schools adhering to a No Excuses philosophy and the decline in student mobility contribute to the improvement in the sector. Even though inclusion of the student mobility and selection variables reduces the magnitude of the estimated No Excuses effect, it remains highly significant in all specifications. Student mobility, largely unstudied in this context, appears to contribute substantially to the improvement of the sector. This finding highlights the importance of patience in understanding the effects of a large-scale reform that opens the education sector to many new entrants of variable quality and that precipitates extensive switching among schools. Finally, although selection into charter schools on the basis of prior achievement and behavior becomes on average more positive over time, there is little evidence that these have been an important component of the improvement in charter school value-added.

We begin with a brief overview of the charter school market in Texas, followed by a description of the Texas Schools Project microdata used in the study. Then we discuss the various approaches used to measure school quality and describe the relative improvement of the charter sector. The final two sections examine the contributions of specific factors to the observed improvements and discuss policy implications and directions for future study.

II. BACKGROUND

A. Texas School Accountability Rating System

Since 1993, Texas has assigned annual ratings to all public schools through an integrated accountability system. On August 1st each year campuses receive a rating of, in order of decreasing distinction, *exemplary*, *recognized*, academically *acceptable*, or academically *unacceptable*.[3]

While the requirements for each rating level have evolved over time, the school-level pass-rate on state-wide standardized exams have been the primary factor determining ratings throughout. In order to achieve a given rating level, school pass-rates in *all* tested subjects must meet the pass-rate standard for that rating. All students in grades 3-10 are tested in math and reading/ELA, while only certain grades are tested in writing, science, and social studies. In addition to the school wide pass-rates on the standardized exams, the system also requires schools to meet the pass rate standard for certain ethnic and income student subgroups. The four subgroups eligible for individual evaluation are economically disadvantaged students, whites, blacks, and Hispanics.[4] Any subject with enough test takers belonging to a given racial subgroup are evaluated on the performance of that subgroup-subject combination in addition to the campus-wide performance in the subject. The number of test takers needed for a specific racial subgroup's performance to count in the rating formula is called the minimum size requirement.[5] Failure to meet the pass rate standard for any exam for any evaluated subgroup can lead to a rating reduction.

---

[3] Through the parallel Alternative Education Accountability (AEA) system, Alternative Education Campuses (AECs) are rated as either academically acceptable or unacceptable. 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 (TEA 2007). During the period of analysis, approximately 5% of rated campuses fall under the AEA system each year. In the remainder of this paper, I focus on schools subject to the standard *exceptional*, *recognized*, *acceptable* and *unacceptable* ratings.

[4] 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, 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.

[5] . The MSR is a deterministic function of the number of test takers at the school. It varies between 30 and 50 students and can differ between subjects within the same school in a given year.

There are three components of the system that allow schools that fail to meet these absolute pass rate standards to move up one rating by meeting certain extra criteria. These are Required Improvement (RI), the Texas Projection Measure (TPM), and exceptions provisions. RI and TPM aim to help schools that fail to meet absolute standards but show test score growth from year to year. The exceptions act to mitigate the fact that larger and more diverse schools are subject to evaluation on a greater number of subgroup and subject combinations.

With few statutory incentives tied to ratings, the effectiveness of the TEA accountability rating system relies on a market response. Negative consequences exist only for the *unacceptable* rating while official rewards are reserved for the *exemplary* rating. Schools that receive an *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 with an *unacceptable* rating can result in escalating sanctions up to and including dissolution. *Exemplary* rated campuses are exempted from certain regulations in the Texas Education Code, such as minimum class size requirements. No distinction is made between *acceptable* and *recognized* schools in terms of awards or punishment.

Additionally, 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.

B. Texas Charter School Program

Since enacting charter school legislation in 1995, the Texas charter sector has grown into one of the largest in the nation. It ranks second nationally in both the number of charters operating and the number of students served by charters in 2010-11.[6] We first discuss the enabling legislation and subsequent modifications and then describe the growth of the Texas charter sector.

---

[6] U.S. Department of Education (2014), Table 216.90 [http://nces.ed.gov/programs/digest/d13/tables/dt13_216.90.asp, accessed June 30, 2014].

1. Institutional Structure

The Texas Education Code establishes four types of charters: home-rule school district charters, independent school district charters, university/college campus or program charters, and open enrollment charters. Open-enrollment charters, which are the focus of this study, constitute the majority of charter schools and educate a substantial fraction of the students enrolled in the sector. Open-enrollment charters are awarded under the auspices of the Texas State Board of Education, which acts as the primary overseer for these schools. These schools are independent public educational entities, and the state designates a unique county-district identifier for schools operating under each open enrollment charter. District charters, by contrast, are established by and accountable to the school districts in which they reside. The small number of university charters makeup the remaining charters in the state and their establishment and operation is similar in character to open-enrollment charters. Thus, we make no distinction between these and open enrollment charters and include university charters in all of the estimates. No home-rule district charters have been established as of this writing.[7]

The defining feature of open-enrollment charter schools is their receipt of public funding without many of the regulatory restrictions, chiefly in the realm of personnel, inherent in traditional public schools. Outside of the requirements imposed by No Child Left Behind legislation for teachers in core areas in any open-enrollment charter receiving federal funds, these charter schools have almost no restrictions on hiring and firing. They may hire teachers who currently lack certification or bring skills and experiences that may not be rewarded in conventional public schools. In addition, open-enrollment charters are able to set salary and benefit schedules freely. By contrast, district charters maintain the hiring and salary rules of their home districts. This distinction leads to some important differences in the characteristics of staff: open-enrollment

[7] Home rule charter districts offer the possibility of increased flexibility for the entire district, but they also have a number of procedural requirements including approval by local voters. The Dallas Independent School District had met the initial requirements and had a charter commission that was developing a charter for the voters, but the commission voted to stop the process in January 2015. See http://www.homerulecommission.com/ [accessed October 31, 2015].

charters tend to employ less experienced teachers who are less likely to have a post-graduate degree than teachers in traditional public schools. Open enrollment charters also pay, on average, lower salaries.

Although district charters offer a degree of parental choice, they involve significantly different incentives for the traditional schools – because total district enrollment and revenue is unaffected by movement to charters. Additionally, because they involve existing personnel, support structures, and general institutional framework, the dynamics of start-up are quite different from those for new open enrollment charters. In some cases, it becomes difficult to distinguish the characteristics of a district charter from those of other schools in the district, and a number of district charter schools were actually stripped of their charter status within the past few years. Therefore, we focus on open-enrollment charters, because they more closely approximate new entrants into a competitive market.

Despite differences in hiring and staffing, all charters in Texas are similar in their stated goals to implement new curricular and disciplinary practices that improve the educational outcomes of their students. The path to achieving these goals differs, however, as both the public mission statements and operational choices of charters vary widely across the sector. For example, many combine standard skills enrichment with an emphasis on discipline; others center the curriculum on more specialized interests such as athletics, the sciences, or music and the arts.

Regardless of curriculum, all charters are subject to the same accountability and testing requirements as traditional public schools. Performance on these achievement measures captures the quality dimension central to the enabling legislation and forms the heart of our evaluation of performance.

Institutionally, there is not a one-to-one match between each charter granted and a specific school (called a campus in Texas). A charter school management organization (CMO) can apply for and hold more than one charter, and each charter can include multiple campuses in the same manner that a traditional public school district can include multiple campuses. As a general rule, each charter applies to one geographic market, and a CMO entering multiple markets will have multiple charters.

Figure 1 illustrates the institutional structure of the Texas charter sector and the dimensions over which a CMO can expand operations using America Can!'s entry and growth through 2011 as an example.

America Can!, a 501(c)(3) non-profit organization, successfully applied for a charter in Dallas and operated one of the first charter schools in Texas in 1997. This CMO subsequently expanded along two dimensions. First, it received an additional four open enrollment charters (covering Houston, San Antonio, Ft. Worth, and Austin) between 1999 and 2005 for a total of five charter districts; and second, it increased the number of campuses operated in three of these charter districts. This pattern highlights a key aspect of the regulatory structure of charter schools in Texas: The approval process of charter districts in good standing to expand the number of schools is far less involved than the process of applying for a new charter, suggesting that the cost of procuring approval for an additional school is likely to be modest relative to other costs associated with adding a school.

From 1997 to 2000, there was no statutory limit on the number of open-enrollment charters granted to management organizations that committed to operate schools that served at least 75 percent "at-risk" students, although the number of unrestricted open-enrollment charters was limited to 100. Two changes were made in 2001. In response to reports of poor performance and mismanagement at some schools, the legislature relaxed the at-risk student composition constraint.[8] At the same time, a strict limit of 215 was imposed on the total number of charters awarded under the open-enrollment program. This limit implicitly advantages existing charter holders by restricting the entry of new charter holders in an environment that permits incumbent CMOs to expand by opening new campuses.

2. Open-Enrollment Charter School Growth

Figure 2 illustrates the growth of open enrollment charters between 1995 and 2011. Prior to 2001, entry of charter school operators and the establishment of new districts constituted the bulk of expansion in the charter sector, as both the number of charter holders and districts increased. After 2001, however, the numbers of charter holders and districts remained roughly stable (around 150 holders and 200 districts), while the number of schools roughly doubled.

[8] Even though the at-risk requirements were modified, the charter sector has continued to enroll an increasingly larger share of poverty students compared to the traditional public school sector.

Figure 3 shows the stock and flow of charters by type. It includes the number of charter districts by active status relative to the state limit as well as the number annual charter authorizations and discontinuations. The number of charters increased through 2001 partly due to the elimination of the separate "at-risk" charter category and the more than doubling of the cap on unrestricted open enrollment charters. The annual increase in the number of new charter districts, however, declined steadily between 1999 and 2002. Exit of charter school operators during the period spanning 2000 – 2011 contributed to these changes as some had their charters revoked and others voluntarily surrendered them. Most of the increase in charter schools, however, can be attributed to expansion of campuses among existing charter districts.

III. DATA

The cornerstone of this research is the microdata constructed by the Texas Schools Project at the University of Texas at Dallas. These data include test scores, demographic characteristics, and information on school attendance and academic programs for a stacked panel of students and schools.[9] Our analysis focuses on over 400 separate charter school campuses and their enrollees for the period spanning 2001 to 2011. School information includes location, grade levels offered, enrollment, charter school type, state accountability rating, and information on all staff. Student information includes demographics, mathematics and reading test results, school attended, grade, and academic program. Students who switch schools, including those who transition between traditional public and charter schools, can be followed as long as they remain within the Texas public school system.[10]

Mathematics and reading assessments come from two statewide criterion-referenced achievement tests that were administered during our period of study. From 1993 - 2003, the Texas Assessment of Academic Skills (TAAS) was administered each spring to eligible students enrolled in grades three through eight. In 2003, Texas introduced a new exam called the Texas Assessment of Knowledge and Skills (TAKS).[11] TAKS expanded the number of subjects for which students were required to demonstrate proficiency and elevated the difficulty of the tests. The tests are not vertically aligned. Thus, these tests cannot be used to measure absolute changes over time in charter school quality; rather they provide information on performance relative to other students and schools in the same grade and year.

Because the test structure, number of questions, and average percent correct vary across time and grades, we standardize all test scores to have a mean of zero and a variance equal to one for each grade and

---

[9] A more detailed description of the underlying database can be found in Kain (2001) and other publications on the website for the Texas Schools Project: http://www.utdallas.edu/research/tsp-erc/.

[10] Private schools enrollment in Texas remains relatively small at less than six percent in 2011 (U.S. Department of Education (2014)).Moreover, in 2010 only 23 percent of people born in Texas had migrated to another state, making it the state with the lowest out-migration rate in the nation(Hanushek, Ruhose, and Woessmann (2015)).

[11] The TAKS exam was recently repealed by the Texas legislature and schools will now transition to End of Course Exams.

year. To address potential concerns associated with imposing a new testing regime, we examine the sensitivity of the results to changes from TAAS to TAKS. We also standardize Spanish language tests separately to avoid potential bias that may arise from pooling.

Any school without students in the TAAS/TAKS data is excluded from the sample; therefore, our number of charters will differ from public records of the number of authorized charter schools.[12] Also omitted are those charter schools exclusively serving children with special needs, residents in treatment programs, or students with diagnosed behavioral problems.

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 remainder of this paper focuses on campus level reenrollment rates when discussing school choice.

For the subsequent analysis of the sources of charter sector improvement, we construct a unique database that incorporates the operational focus of each charter school. Specifically, on the basis of information gathered through interviews and records investigations, we classified each CMO on the basis of whether or not it adheres to a No Excuses philosophy as defined below in Section 6 and in Appendix A. We also use the components of the "No Excuses" taxonomy in separate analyses.

I also make use of publicly available campus-level data can be downloaded from the TEA Academic Excellence Indicator System and Accountability System website. 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.

[12] Note, however, that students do not have to have to complete the tests to be included in the TAAS/TAKS file.

Demographic characteristics of schools with different ratings are shown in Table 1. The standard rating system covers 97% of students in Texas over the time period with 45% of students enrolled at a school rated *acceptable* while only 3% were enrolled at schools rated *unacceptable*. Schools with higher ratings tend to be smaller, they have a higher proportion of white students, and a lower proportion of black, Hispanic and low income students. Table 2 shows pass-rates gains by the school's year-to-year rating transition. For example, schools that go from recognized to exceptional from one year to the next showed on average a 2.73 percentage point increase in their math pass rate between the two years. Looking at the total column on the right of both panels, you can see that the mean pass-rate gains for math and reading are monotonically decreasing in the prior year's rating. Note, however that the difference in the gain between ratings is largest between *unacceptable* and *acceptable* suggesting that either the information signal at is most informative at this boundary or that the direct pressure from sanctions, which only occur at this margin, increase the incentive to improve performance.

Distributions of math pass rates and reenrollment rates by prior rating are presented in figures 4 and 5. In figure 4 we see that the distributions of math pass rate gains are more dispersed but with a higher mean as rating decreases. Panels A in figure 5 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 of value-added to math test scores. While suggestive, the differences between these distributions in all of these figures fail to address the fact that schools at different rating levels likely differ in unobservable ways related to the quality of the education they provide. The following section lays out my approach to addressing these challenges.

IV. EMPIRICAL METHODS FOR INDENTIFYING THE EFFECTS OF SCHOOL RATINGS

A. Conceptual Framework

This section outlines a framework for thinking about the relationship between school ratings, and subsequent school performance and the family school choice decision. For simplicity, assume that schools are assigned either a high or low rating with schools rated low being subject to sanctions from the ratings authority.

The resulting accountability pressure from the receipt of a low rating come directly from sanctions and indirectly from any market response related to the information signal about school quality embodied in the rating. Direct pressure comes from the loss of independence after state intervention and the threat of escalating sanctions if a low rating is received in subsequent years. These interventions are designed to improve academic achievement and take the form of interventions outside the purview of local school administrators. Immediate sanctions include posting a public notice of deficiency or the appointment of a special campus intervention team, while consecutive years of a low rating can lead to reconstitution or dissolution. School administrators are assumed to try to avoid sanctions in favor of maintaining control over their own schools and therefore prefer not to receive the low rating. The escalating nature of sanctions provide extra incentive to improve test scores for low rated schools.

Indirect pressure through market forces comes from reputational effects. Accountability ratings act as a noisy yet low cost source of information about school quality to families, administrators, higher level state and federal agencies, and teachers. The information is particularly important when other information about quality is costly to obtain. School administrators are assumed to prefer a favorable reputation for several reasons including personal career motives, to attract higher quality teachers, and, especially for charter schools, to maintain enrollment and the commensurate state funding. Combined, the direct effect of external sanctions and the indirect reputational effect act to unambiguously increase the incentive to improve performance for schools that receive the low rating compared to schools that receive the high rating.

The framework for thinking about the effect of rating information on the family school choice decision focuses on expectations about academic achievement. As with the school performance framework, ratings affect a school's reputation indirectly through the information signal about quality, and directly from changes to a school's expected academic performance due to interventions. All else equal, a higher rating signals better academics and makes a school more attractive. On the other hand, the sanctions and interventions associated with the low rating may be expected to produce greater achievement gains relative to what would be realized if the incumbent administration were to remain in control. If interventions are effective, then rational families will revise their expectations about school quality upward when interventions are triggered.

In this way, the imposition of sanctions positively affects the likelihood of enrollment as families choosing between similar schools may prefer a school facing an intervention to a similar school maintaining the status quo. The reputational effect of a low rating should negatively affect the likelihood of enrollment as the lower rating serves as an information signal about the quality of the school. The overall effect of receiving a lower rating and the associated interventions is therefore ambiguous. It is important to note, that only the indirect effect remains if there is no intervention triggered by the lower rating. In this case, if families receive new information from ratings and act on the information, then a lower rating will result in lower rates of reenrollment.

All of these effects are expected to be larger at charter schools compared to TPS for two reasons. As public schools of choice, competitive forces in the charter school sector act to increase the incentives to increase test scores. Further, with all public funding determined by enrollment charters also have strong incentives to maintain a high quality expectation and demand for seats at their schools. Second is the fact that charter school students all have the outside option of returning to their neighborhood TPS while TPS students may face geographic or capacity constraints when trying to switch to another TPS or a charter school. Because of this, charter families have a greater ability to quickly respond to a ratings release than families attending a TPS. [13] Combined with the fact that charter families have already demonstrated knowledge of their school choice options[14], these factors support the evaluation of charters separately from TPS.

The relative strength of these effects for potential enrollees to a specific school can differ between student type: those currently enrolled at the school (internal) and those not currently enrolled (external). The internal students decide whether or not to reenroll while the external students decide whether or not to transfer to the school. For the internal student, the marginal contribution of the rating as a signal about school quality is likely smaller than the effect for external students because internal students have direct family experience and are more likely to have a social network as a source of additional information about a school. The remainder of this paper focuses on internal students and uses the reenrollment likelihood as the metric of school choice. Early work on another project looks at how the profile of incoming students and the characteristics of the schools from which the separate differ by school rating.[15]

B. Identification Issues

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 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 differences in school feeder systems, administrative changes, prior achievement, and neighborhood transitions.

---

[13] Given the late July or early August rating release schedule, the one-year ahead corresponds to a 1-2 month response window, while the two-years ahead corresponds to a 13-14 month response window.
[14] Hanushek et al (2007) find that charter school students in Texas exhibit greater mobility than traditional public school students.
[15] See appendix E for descriptive statistics in this area.

This paper takes two approaches to address these problems. The first 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 school for the subset of schools that receive at least two different ratings during the sample period. 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 second 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 treat 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 very specific subsets of schools based on a very specific source of rating variation. The remainder of this section describes the details of these two approaches to identification.

C. School Fixed Effects

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. This FE strategy compares the outcomes from each rating in a single regression and controls for unobservable characteristics of schools that differ between schools but are time invariant within a school. The acceptable rating is excluded and serves as the reference rating.

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

V. $$RR_{it} = \sum_{A=\{E,R,U\}} \theta^A A_{it-1} + X_{it}\beta + \tau_t + \gamma_i + \varepsilon_{it}, \quad (1)$$

where RR_{it} is the reenrollment rate among eligible students at campus i in year t. The θ coefficients on the lagged accountability rating dummies give the mean difference in the reenrollment rate for schools that received *exemplary*, *recognized* and *unacceptable* ratings in the prior year relative to *acceptable* schools. I control for school specific time varying characteristics X_{it} in addition to school and year dummies, γ_i and τ_y.

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

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

where PR_{it} is the pass-rate for school i in year t. The θ 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]

I also estimate a gains model that is a special case of equation (2) where the persistence parameter λ from the lagged dependent variable model is restricted to be one:

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

In all likelihood, the persistence of the campus pass-rate likely takes on a value less than one rendering the gains model invalid. Imberman (2011) proposes that in conjunction with the IV estimates from equation

[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.

(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. However, administrative and neighborhood changes over time that have persistent achievement effects make it difficult to assert causality as some of these changes may be directly influence ratings and future performance. The RD approach in the next subsection address these concerns.

D. Fuzzy Regression Discontinuity Design

A school can be dropped down a rating level if just one subgroup that is large enough to be evaluated fails to meet the pass rate requirement. This results in an ex-post campus-level probability of receiving each rating that depends discontinuously on both the size and the performance of each subject in each subgroup combination. The fuzzy RD approach in this paper uses these discontinuities in rating assignments to implicitly compare 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 school demographic group sizes. 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, *recognized* and *acceptable* without loss of generality. This approach then can be applied separately at each of the other two rating margins: *exemplary-recognized* and *acceptable-unacceptable*.

The first running variable, across the pass rate discontinuity, is defined as the lowest pass rate in any exam among all student groups that are large enough to count in the accountability rating formula. Across the five subjects, if a school has one large demographic subgroup that fails to meet the pass rate required for *recognized*, that school will face an increased probability of receiving the lower *acceptable* rating. However, once the worst performing large group passes the required pass rate, the probability of receiving the *recognized* ratings increases discontinuously. Panels (a) and (b) of figure 6 illustrate campuses that have values of this running variable on either side of the discontinuity, represented by the horizontal axis. [17]

The second running variable, across the group size discontinuity, is defined as the largest group size among all subjects that fails to meet the pass rate standard for *recognized*. If a school has one underperforming subgroup that is large enough to be evaluated, that school will face a high probability of being rated acceptable instead of acceptable. However, once the largest of the underperforming subgroups becomes too small to be evaluated then the probability of being rated *recognized* increases discontinuously. Kane & Staiger (2002) discuss this discontinuity as it relates to the fairness of school accountability systems. Panels (c) and (d) of figure 6 illustrate campuses that have values of this running variable on either side of the discontinuity, represented by the vertical axis.

These discontinuities can not be combined into a single running variable due to fundamentally different units of measure between variables.[18] This fact necessitates the use of a multidimensional RD design. Reardon & Robinson (2012) outline several ways to implement a regression discontinuity design with multiple running variables. To identify the effect of the assigned rating, I use a combination of what they call a *binding score RD* and the *fuzzy frontier 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.

Figure 6 shows the interrelatedness of the two dimensions of treatment graphically. Schools are treated with a higher probability of being rated acceptable if they have a student subgroup in quadrant IV. The pass rate discontinuity looks at schools that are one either side of the boundary between quadrants I and IV, while the group size discontinuity looks at schools on either side of the boundary between quadrants

[17] Appendix C lists the pass-rates that a campus must achieve for all evaluated subgroup-subjects to achieve each rating. Since these differ by subject and over time, I center all pass rates on the requirement for that subject in that year for that rating.

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

III and IV. The relevance of each indicator depends on where it lies in this 'pass-rate' – 'subgroup size' space. Negative values for the pass-rate on the y-axis *and* positive values for the subgroup size on the x-axis are needed for treatment. If either of these conditions is not met, then that subgroup does not induce treatment. Small changes in the size or performance of the marginal subgroup can result in a campus being moved into or out of treatment.

I analyze this fuzzy RD design through an instrumental variables methodology:

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

where y_{it} is the outcome for school i in year t. This is regressed against a dummy indicating if school i receives an *acceptable* rating or lower in the prior year, A_{it-1}, and flexible function $f(\cdot)$ of the running variable D_{it}. X_{it} is a vector of student and campus demographics, and τ_t represents a set of year dummies. While not strictly necessary, the covariates and year fixed effects could be included to improve precision. In the specifications I report, I do not control for these other factors, which I show to be continuous across the thresholds in Appendix B.

To instrument for the endogenous rating variable A_{it-1}, I use the treatment dummy, T_{it-1} which is equal to 1 if the school has a subgroup in quadrant IV in figure 6 and is equal to 0 otherwise. This can be decomposed into the reduced form:

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

and first stage:

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

The reduced form coefficient θ^{RF} gives the effect of having a low performing group that counts, regardless of if it results in an *acceptable* rating. The first stage coefficient θ^{FS} measures the increase in the probability of receiving an *acceptable* rating when a campus is treated with a large underperforming group.

For the running variable control function $f(\cdot)$ I use a non-parametric local linear regression (LLR). The identifying assumption that T_{it} only affects y_{it} through its effect on A_{it-1} is achieved by limiting the sample to schools which lie within a small bandwidth of the discontinuity. In all analyses, I use the optimal bandwidth determined using the data driven algorithm set out in Imbens and Kalyanaraman (2012).[19]

Concerns about the ability of schools to differentially sort out of treatment is mitigated by the fact that treatment status is only realized on the testing day when the number of test-takers is determined and the exams are given. 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. 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. These data are not available for all years however, limiting the possibility of testing this assumption.

[19] For implementation, I use the user written Stata command rd (Nichols 2011).

V. ESTIMATES OF THE EFFECTS OF SCHOOL RATINGS

A. School Fixed Effects

I measure achievement in terms of the school level pass-rate on the state administered math and reading exams that are used to construct the ratings. These are publicly available data and capture the most salient metric of a school's quality. For the school choice analysis, I use the school level reenrollment rate among students for whom the subsequent grade is offered (eligible students).

The relationship between ratings and the next year's campus-wide math pass-rate are given in table 3. The first two columns give the coefficients on the prior rating dummies in equation (2). To address the violations of conditional strict exogeneity in this equation, columns (3) and (4) instruments for the lagged pass rate in equation (2) using the twice lagged pass rate. Gains models in columns (5) and (6) correspond to equation (3). Even columns include school fixed effects to control for time-invariant differences between schools. When FE are included estimates are identified by within school variation in ratings and pass rates. Across all FE specifications, there are large and significant differences between rating levels in subsequent performance.

In addition to the causal effect of the rating, other factors may potentially be contributing to these large differences. First are persistent changes being implemented at the school level that affect school quality. Persistent changes in quality will contribute both to the prior rating and to this year's performance. A new principal, for example, may implement curricular or hiring changes that result in sustained pass rate changes. To the extent that such administrative changes are caused by a low rating, these changes in true quality are the effects of ratings that the FE approach intends to measure. Second are community or neighborhood demographic trends that are out of the direct purview of the school. Neighborhood transitions may contribute to pass rate gains and rating improvement, and the improved school district may then re-enforce the demographic trend. These demographic trends may also be reflected in the FE estimates. Third, and most worrisome, is the likely presence of mean reversion.

The inclusion of lagged achievement and demographic controls help mitigate the contributions of the first two concerns. In an attempt to speak to mean reversion, I estimate a model that instruments for the lagged pass rates with the twice lagged pass rates. With the uncertainty over the degree to which mean reversion drives the estimates, I take the conservative approach and focus on the IV estimates in column (4). These estimates indicate that the largest difference between adjacent ratings occurs between *acceptable* and *unacceptable*. Approximately half of the pass rate difference between exemplary and unacceptable schools, about 4.5 percentage points, occurs at this boundary. Results for reading in table 4 show virtually no effect of the different ratings other than in the gains model. It is not out of the ordinary in the education literature to find effects on math scores and smaller or no effect on reading scores.[20]

The inclusion of time varying achievement and demographic characteristics may partially mitigate the first two confounding factors. However, there remains the serious concern that mean reversion may still be playing a part. I attempt to fully addresses all three of these concerns using an RD approach in the next section. However, before turning to the RD results, I will use the FE approach to evaluate the effects on reenrollment where mean reversion is not an issue

Table 5 gives results of the reenrollment regressions including school fixed effects. Columns (1) and (2) have the reenrollment rate among eligible students the year after the rating as the dependent variable. It is not surprising that there is no effect due to the short 2-month window of time between the release of ratings and the start of the next school year. In columns (3) and (4) the dependent variable is the reenrollment rate among eligible students two years after the rating. With a 14 month response window there is a larger and significant, though still small 2.8 percentage point lower reenrollment rate when a school is rated *unacceptable* compared to when it is rated *exemplary* in column (4).[21] Again the largest

[20] See Dee and Jacob (2009) and Wong, Cook and Steiner (2009) for examples.
[21] Estimates from Hanushek, Kain and Rivkin (2004) regarding the effects of student mobility on overall campus achievement suggests that this 2.8 percentage point difference would translate to a small 0.033 standard deviation change in test scores. 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. This 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.

difference between neighboring ratings occurs at the *acceptable-unacceptable* boundary. These reenrollment effects are smaller than estimates without including school fixed fixed effects. This suggests that mean differences between schools over time correlated with ratings are predictive of differences in student turnover. This is not surprising as Hanushek, Kain, and Rivkin (2004) have documented the negative relationship between turnover and achievement. Nevertheless, it is not possible to break down if the lack of a reenrollment effect is due to little new information being learned from ratings or from the high cost of switching schools.

While the school reenrollment estimates are not subject to the same measurement error as the pass rate estimates, they do still combine the rating information effects with changes in the true school quality and any neighborhood transition. Now I move to the RD analysis in the next section where I aim to address all of these concerns.

B. Fuzzy Regression Discontinuity

As described in section 5, I use two different running variables in the RD analysis. Combined with the three different rating boundaries, this yields six quasi-experiments for each outcome. Each specific outcome variable in the results sections has its own specific first-stage F-statistic for each of the six RDs due to different subsets of schools falling within the bandwidths of the size discontinuities. For that reason, I will present the first stage effects of graphically for only the estimates of the effect on math pass rates. First stage magnitudes are similar across the other outcomes.

Figure 7 illustrates the first-stage effects of having a large group that cross the performance cutoff for each rating. In these figures, I plot the proportion of schools rated below the specified rating boundary on the vertical axis against the number of percentage points *below* the pass-rate requirement for the lowest performing large group on the horizontal. In this way, treatment with a higher likelihood of a lower rating occurs to the right of the discontinuity. Panel (c) for example gives the proportion of schools rated below *exemplary* by the distance in percentage points from the *exemplary* pass-rate requirement for the lowest performing large subgroup at each campus. For schools where this group just fails to meet the required pass-rate for *exemplary* the proportion of schools rated below *exemplary* increases over 40 percentage points relative to schools that do meet the requirement. Similarly, in panel (b) we can see that the proportion of schools that miss out on the *recognized* rating increases over 55 percentage points once one large group fails to meet the standard for the *recognized* rating. One specific aspect of the system is clearly visible in panels (a) and (b). The *Required Improvement* provision allows for schools that should be rated *acceptable* or *unacceptable* but are within five percentage points of the standard for the next higher rating to be moved up one rating if they have shown evidence of past performance gains that would lead to meeting the standard within two years. For this reason, these plots exhibit two discontinuities, one at the standard, and a smaller one at 5 points below the standard (here to the right of 0). I do not evaluate any effects from the second discontinuity.

The discontinuity in panel (a) is much smaller at under 10 percentage points. This is due to the other aspects of the accountability system outlined in section 3.2 and in Appendix C which tend to be more consequential for campuses near the *unacceptable* rating boundary, notably the exceptions provisions. For this reason, the instrument at this boundary is weakest.

Figure 8 illustrates the first-stage effects at each rating boundary using the group size running variable. Again, the vertical axes are the proportion of schools rated below the stated boundary, but now the horizontal axis is the number of students over the size requirement of the largest low performing group. Panel (b), for example, gives the proportion of schools rated *acceptable* or below by the number of students from the size requirement for the largest group at that campus with a pass rate below the requirement for *recognized*. For schools whose group is just large enough to be evaluated, the proportion of schools rated below *recognized* jumps approximately 70 percentage points. Similarly, in panel (c) we can see that the proportion of schools that miss out on the *exemplary* rating also increases approximately 70 percentage points across the boundary and in panel (a) the proportion rated *unacceptable* increases about 25 percentage points across the cut-off.

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 highly statistically significant. Having established the strength of the first stage in general, I know move on to the RD estimates of the effects of ratings on pass- and reenrollment rates. In tables 6 and 7 the first three columns give the estimates at the three rating boundaries using the group size discontinuity and the second three columns give estimates using the pass rate discontinuity. Columns labeled "E to R" measure effects at the *exemplary-recognized* boundary, "R to A" corresponds to the *recognized-acceptable* boundary, and A to U is the *acceptable-unacceptable* boundary. All models control for the distance from the pass rate using a non-parametric local linear regression.

Table 6 summarizes estimates of the effect of receiving the lower of two ratings on a school's math and reading pass-rates the next year. What is clear from these estimates is that a negative shock at the *acceptable-unacceptable* boundary results in greater math pass rate 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. The IV estimates in the *acceptable-unacceptable* columns are similar using either instrument and indicate a 7-9 percentage point gain in math for schools rated *unacceptable* compared to similar schools that are assigned the *acceptable* rating. The effects at the other boundaries are generally insignificant. These effects can be seen in figure 9, which plots the reduced form relationship between mean campus math pass-rate and the distance to the three rating discontinuities for both instruments. There is no discontinuity across the border in panels (a)-(d). With the exception of a large but highly insignificant effect at the acceptable-unacceptable boundary using the pass-rate instrument, there is no again no estimated effect on reading gains.

Compared to the FE estimates in the prior section, the RD estimated effects are even more concentrated at the low end of the rating distribution and larger than the FE estimates. Since the RD estimates represent the pure reputational effect felt by school administrators from communities and superiors, the FE differences between the higher rating levels indicates that confounding factors may be playing a role at these boundaries. To the extent that these factors also exist at school near the *acceptable-unacceptable* boundary, this suggests that schools that are experiencing adverse community or administrative changes are more likely to receive the unacceptable rating.

Panel A in table 7 presents the reenrollment results looking at school rosters the year after the rating release, and panel B looks at reenrollment two years after the rating release. In general estimates are very imprecise, but in most cases, the point estimates of the effect in two years is larger in magnitude than the immediate effect. This corroborates the FE estimates which show that the additional time from the release of ratings allows for more possibilities for families to explore outside schooling options. Most point estimates are negative or zero, lo response or a slightly higher reenrollment rate at higher rated schools. Again, this agrees with the pattern seen in the school FE estimates. However almost all estimates are insignificant. Since the RD estimates capture the pure effect or information while holding quality constant, the lack of a significant response suggests one of two things. Either families' expectations about school quality are unaffected by the empty information signal embodied by the assigned ratings, or there is little new information learned from ratings.

Alternatively, consider the case where parents have no information about school quality prior to rating assignment. If this were the case, one would expect the FE and RD estimates to be more comparable as the confounding effects of dynamic changes to school quality would not affect families' enrollment decisions in the FE model other than through the assigned rating. The qualitative difference between the two models therefore supports the notion that families do have substantive prior information regarding school quality.

It can not be overemphasized that the RD estimates in this section represent LATEs and are not necessarily indicative of how schools and families away from these margins would respond to a rating change. Nevertheless, this identification strategy involves boundaries at the low, middle, and high points of the school quality distribution allowing for range of potential LATE estimates. Further, with the two simultaneous running variables, this yields ranges of potential LATE estimates along two different hyperplanes in the two dimensional treatment space.

C. Charter School Subset

The charter school sector in Texas is growing, with over 500 schools educating more than 5% of the student population as of 2011. Charter schools are public schools of choice that face fewer regulations than traditional public schools especially in the realm of hiring. To ensure that charter schools are effective, they are subject to the same accountability system as traditional schools. As schools of choice, enrollees must apply to a charter, and state funding for these schools is calculated based on enrollment. It is possible that the lack of a clear reenrollment response among TPS families may be related to a low availability of school choice options. Further, with greater incentives to perform, charters may also show a greater pass rate response then TPS. To assess these hypotheses, I repeat the analysis from sections A and B above for only only charter schools, whose administrators face different objectives than their TPS counterparts and whose students all have a neighborhood TPS at which they are entitled to enroll if they so choose. Results are in tables 8-12. The Gains and OLS models show the same pattern as the TPS estimates and are indeed larger in magnitude. The preferred school fixed effects models of math achievement, the IV estimates in column (4) of table 8, strangely show the opposite pattern. These estimates are very imprecisely estimated however, making inference difficult. RD estimates in table 11a show that math pass rates have a much stronger response to rating differences, though the estimates are generally not significant. One puzzling results in need of further investigation is the large negative effect on both math and reading at the recognized-acceptable boundary using the group size instrument. Nevertheless, the much larger point estimates for charters compared to TPS are again consistent there being stronger incentives to perform in the charter sector.

School choice models in table 10 show that charter families show similar response to rating information if we measure reenrollment rates one- or two years ahead. The overall difference between *exemplary-unacceptable* charter schools is similar to that of TPS two years ahead. These are consistent with the notion that TPS students have a greater average cost of executing a quick switch as they 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. While the overall difference is similar, for charters the entire difference is concentrated between the *acceptable* and *unacceptable* ratings. Rating information is having less of an effect for charter families when it comes to those enrolled at schools at the top of the achievement distribution, but a stronger effect on families' decisions to leave the lowest performing schools. The RD results for charter reenrollment vary widely across rating boundaries and between instruments. More investigation is needed here to determine if any meaningful conclusions can be drawn.

D. Testing the Assumptions of RD

For the fuzzy RD estimates to take on a causal interpretation, the position of a campus relative to the discontinuities, the instruments, must satisfy the relevant 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 would be indicative of differential sorting which would violate the exclusion restriction. Figures in Appendix B illustrate the continuity of the mean campus characteristics and densities at the boundaries. There is no heaping on either side of the boundaries in the densities which would be indicative of the possibility that schools may be able to select into or out of treatment. Any discontinuities at the boundaries in plots of school characteristics, demographics and quality across the treatment boundaries would also threaten the causal interpretation of the RD estimates. These figures, and point estimates of discontinuities (not shown) also provide no significant evidence that campuses differ discontinuously along any of these dimensions.[22]

Another important facet of RD research is the bandwidth specification. I use the optimal bandwidth procedure outlined in Imbens and Kalyanaraman (2012) and I control non-parametrically using local liner regression within the bandwidth. LLR is preferred to kernel estimation, especially when using the pass-

[22] School characteristics evaluated include the proportion of students that are white, black, Hispanic, or low income, campus size, the proportion schools that are charters, grades offered, and the number of years in operation.

rate instrument, where 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 main results.

VI. MEASURING CHARTER SCHOOL QUALITY

The primary concern in measuring charter school performance is that unobserved differences between charter school and traditional public school attendees contaminate comparisons of achievement in the two sectors. This is particularly salient in this analysis, as evidence below illustrates the increasingly positive selection of charter school entrants in terms of prior achievement. We begin with a school-level value-added model (which becomes the base specification in our subsequent estimation). In the context of this model, we highlight potential problems introduced by purposeful sorting of students into schools. From that, we consider common alternative approaches to mitigate these problems in the estimation of charter school effects and how these interact with our focus on the estimation of changes over time in charter school effectiveness.

A. School Value-added Model

Equation (1) presents a value-added model in which achievement A for student i in grade g and school s is modeled as a function of prior achievement, prior behavior (D), contemporaneous student and family factors (X), a school fixed effect that is our measure of school quality (δ_s), and a random error:

$$A_{is} = f\left(A_{i,t-1}\right) + 1[D_{i,t-1}] + X_{is}\beta + \delta_s + \epsilon_{is} \qquad (1)$$

Following the literature, we control for prior achievement with cubic functions in both mathematics and reading scores. We also include an indicator for receipt of a disciplinary infraction in the prior year. The vector X includes indicators for race, ethnicity, gender, and mobility. Note that we suppress grade fixed effects and the time subscript to simplify presentation.

School value-added estimates are produced from separate regressions for each year, and the grade fixed effect is a year-by-grade error component intended to capture grade-specific changes over time in the test instrument and any state-level changes in policy. Within this framework, we can estimate the full distribution of school quality across both traditional and charter schools. Further, and key to this study, we can trace the evolution of quality across time and can then consider how market dynamics enter.

The validity of the school fixed effects as measures of quality depends upon the assumption that the prior test scores, disciplinary infraction measures, mobility controls, and other included variables account for confounding factors related to school quality. While a vigorous debate continues about the estimation and use of teacher value-added measures, much less attention has gone into such estimation at the school level. Two concerns have dominated the discussion of teacher value-added but are much less important here. First, researchers continue to debate the extent to which systematic student sorting, both within and between schools, contaminates estimates of teacher value-added. Rothstein (2010) provides evidence of bias introduced by endogenous sorting into classrooms, but Chetty, Friedman, and Rockoff (2014a) find that including one-year lagged achievement along with common demographic characteristics effectively eliminates bias.[23] Our focus on average school quality rather than the effectiveness of individual teachers, however, reduces the relevance of issues related to classroom placement.[24] Second, concerns about the variance of estimation error and the instability of teacher effects, particularly in proposed uses for personnel decisions, have been extensively discussed. These problems are, however, largely related to small samples for individual classroom teachers (McCaffrey et al., 2009) and are much less important at the school level.

Nevertheless, the possibility that these included variables fail to account fully for sorting among schools is still present. Research on charter schools has adopted a variety of approaches to account for unobserved heterogeneity, and the merits of each have now been examined extensively. We focus primarily on school value-added measures and place them within the context of alternative approaches.

---

[23] In unpublished papers, Rothstein (2014) asserts that bias in estimation remains, while in response Chetty, Friedman, and Rockoff (2014b) reject his test.

[24] It may be that classroom placement of students is productive, i.e., average student gains are higher in schools where student groupings and matches with teachers are optimal. For our analysis this is simply reflected in the overall school value-added, and we make no attempt to disentangle such sources of any differences in school value-added.

In terms of internal validity, admissions lotteries constitute the gold standard, as they effectively randomize assignment to charters and in the absence of nonrandom attrition produce consistent estimates of charter school effects. However, only oversubscribed schools conduct admissions lotteries, and an analysis of sector dynamics must cover all charter schools. Lottery studies are still relevant to the extent that they provide evidence on the performance of other estimation methods. Although comparisons between lottery and observational estimates of charter school quality employing value-added approaches do not exist for Texas, Abdulkadiroğlu et al. (2011), Dobbie and Fryer (2013), and Deming (2014) present evidence that lottery approaches and alternative observational identification strategies generate broadly similar estimates in their work on Massachusetts, New York, and North Carolina, respectively.

Matching of charter school students with observationally equivalent students in the traditional public schools from which new charter entrants originate has been used in used in several recent studies; e.g., see CREDO (2013) and Angrist, Pathak, and Walters (2013) along with studies investigating the correspondence of lottery and observational estimates. Although these matching approaches do not address selection on unobservables, they do account for systematic differences in observed characteristics and the composition of traditional public schools previously attended by charter school students.[25] In an evaluation of alternative approaches, Fortson et al. (2012) find that such matching methods produce estimates that are not significantly different from lottery-based estimates over the same sample of schools; estimates produced by regression adjustments without matching of students tend to be fairly close in magnitude though statistically different.

Importantly, while prior discussions have emphasized the degree to which the estimator accounts for unobserved heterogeneity, consideration of the statistical methods also conflates issues related to the

[25] Student fixed effects provides another alternative approach to the identification of charter and traditional public school quality, as each student acts as his or her own control; see Bifulco and Ladd (2006), Sass (2006), Booker et al. (2007), and Hanushek et al. (2007). However, in models with student fixed effect only students who attended schools in both sectors contribute to identification. Estimates based only on switchers may be particularly prone to biases introduced by time-varying student shocks. Moreover, in their study of variation in teacher value-added estimates, Guarino, Reckase, and Wooldridge (2015) find that the types of shocks typically considered problematic in this context appear to introduce less bias into value-added estimates produced by the lagged-achievement model than those produced by other models, including those with student fixed effects.

comparison group for charter schools. This complicates the interpretation of any differences among lottery-based, matching, and simple regression-adjusted value-added models. The general value-added estimator weights each traditional public school on the basis of enrollment, while the matching models weight on the basis of the distribution of traditional public schools previously attended by charter school students and the lottery-based estimators weight on the basis of the distribution of traditional public schools attended by lottery losers. The traditional public school comparison groups for the matching and the lottery-based estimates invariably reflect a geographic distribution similar to that of the charter schools, while the comparison group for the regression model generally reflects the statewide distribution of traditional public school students. This difference may contribute to the finding in Fortson (2012) of a greater similarity between lottery-based and matching estimators.

In our context of dynamic market adjustments with a rapidly growing charter sector, matching model estimates of changes over time in the charter-traditional public school quality differential will reflect any changes over time in the quality of the traditional public schools that students leave to enroll in charters schools.[26] Consider both the response to a change in the quality of a charter school and the expansion of the charter sector. If a charter school improves it is likely to appeal to students from higher-quality traditional public schools. In addition, because a CMO is likely to consider traditional public school quality and the local demand for charter schools in the determination of where to open a school, a decline in the quality of traditional public schools may elevate the probability that a charter school opens in a local community. Each of these processes may lead estimates of changes in charter school quality based on matching methods to diverge from those produced by a statewide comparison, because changes in the pattern of transitions to a charter school would have a negligible effect on traditional public school enrollment shares.[27]

[26] Changes over time in the composition of schools that hold lotteries also changes the control group of traditional public schools.

[27] The findings in Gleason et al. (2010) illustrate the possibility that changes over time in the distribution of traditional public schools can alter estimates of charter school effects. First, the lottery-based method generates substantial heterogeneity in estimated charter school effects. Second, the estimated effect of charter school attendance is much higher for low-income students. This finding is consistent with the possibility that the gains from charter school attendance are likely higher in areas with lower-quality traditional public schools (assuming that school quality tends to be lower as poverty increases). Some of the observed variation almost certainly reflects

Relatedly, any competitive effects of charter schools on the quality of instruction in the traditional sector are likely to be strongest in schools most directly affected by charter school competition. Therefore, matching models might also be more prone to general equilibrium effects that dampen estimates of charter sector improvement. In a preliminary analysis not reported, we found a strong positive relationship between charter school quality and the quality of the origin schools in the traditional sector after controlling for school fixed effects. While this association does not provide causal evidence of a competitive effect, it is consistent with the possibility that such competitive effects may be present.

All in all, the matching and lottery estimates differ from the statewide value-added models in the construction of the traditional public school comparison group and likely also in the sensitivity to the influences of unobserved heterogeneity and traditional public school responsiveness to competition from the charter sector. Consequently, we provide two sets of estimates to illuminate the sensitivity of the findings to the empirical specification. The baseline results use estimates of school quality produced by regression-adjusted value-added models, referred to as statewide comparisons. The second set of results use estimates of school quality produced by matching models similar in spirit to the approach proposed in Angrist, Pathak, and Walters (2013).

Specifically, Angrist, Pathak, and Walters (2013) construct estimates of charter school effectiveness by comparing the achievement of charter school enrollees with a control sample of traditional public school students that fall within the same baseline year- gender-race-school cell. This specification is estimated on observational data and controls for initial achievement and the number of years a student is enrolled in any charter school. They use this estimator to study overall effects of attending charter schools, on average, across a student's career. Because our goal is to estimate the evolution of charter quality over time, we modify their approach. Our year–over–year estimates instead compare the achievement of charter school enrollees to their public school counterparts that fall within the same race–gender-school cell

heterogeneity in charter school effects, but the pattern is consistent with the existence of heterogeneity in traditional public school quality as well.

controlling for a flexible function of past achievement scores and other observable characteristics. Hence, our matching estimates provide a more ``localized'' estimate of changes in school effectiveness.

VII. EVOLUTION OF THE CHARTER SCHOOL QUALITY DISTRIBUTION

In this section, we describe changes over time in charter school mathematics and reading value-added between 2001 and 2011 relative to traditional public schools and subsequently examine the contributions of school improvement, school closures, and the entry of new schools to these changes. We adjust for differences in years of operation across school by regressing the school quality estimates on a set of indicators for the first, second, third, and fourth years of operation and then use the residuals as the measures of quality, though unadjusted estimates (not reported) reveal almost identical patterns. We provide parallel estimates for a statewide comparison group and for the more localized comparison group provided by the matching estimator.

A. Statewide Comparison

We first describe changes over time in the distribution of charter school value-added in comparison to traditional public schools across the state. Then we consider the contributions of school improvements, closures and entry to these changes.

1. Performance trends over time

Figure 10 illustrates changes over time in the 25th, 50th, and 75th percentiles of charter school value-added in mathematics and reading relative to the corresponding percentiles of the traditional public school mathematics and reading distributions (The full distributions are shown in Appendix D). Relative improvements in charter school mathematics value-added (Panel A) occurred throughout the distribution and decade, though increases at the 25th percentile were roughly twice as large as those at the 50th and 75th percentiles.[28] In 2001, charter school mathematics value-added was roughly 0.3 standard deviations

[28]Note that the differential declines between 2001 and 2003, the period in which the state switched from the TAAS to the TAKS test, and between 2003 and 2011 when the TAKS was used throughout. This consistency indicates that the observed pattern is not just a testing phenomenon.

below traditional public school value-added at the 25th percentile, but by 2011 that gap fell to less than 0.1 standard deviations. By comparison, the difference at the median declined from slightly more than 0.1 standard deviations to roughly zero, and the difference at the 75th percentile changed from a charter school deficit of almost 0.1 standard deviations to a charter school advantage of roughly 0.05 standard deviations.

Panel B illustrates charter sector improvement in reading value-added. At the 25th percentile the gains are comparable to those for mathematics; the reading value-added differential went from a charter school deficit of almost 0.25 standard deviations to a slight charter school advantage by the end of the period. At the 50th and 75th percentiles charter schools start very close to traditional public schools, but they also show far smaller improvements.

Overall, the average performance of charter schools relative to traditional public schools improved over this period by 0.12 standard deviations in mathematics and 0.09 standard deviations in reading. Kolmogorov – Smirnov tests for equality of the charter and TPS mathematics and reading distributions at the end of the observation period reject the equality of the distributions at the 1 percent level. These increases are similar in magnitude to typical estimates of a one standard deviation difference in teacher value-added or the benefits of a substantial reduction in class size. The smaller change for reading is consistent with the frequent finding that schools exert a greater effect on learning in mathematics than in reading.

In addition to the average improvement there is also a significant fall in the variance of charter school quality. While charter school quality is much more dispersed than TPS quality at the beginning of the period, the improvements – particularly at the bottom end – draw in the long left tail of the charter school quality distribution over time (see Appendix D). By the end of the period, the overall quality distributions, especially in reading, look very similar.

Any interpretation of these figures in terms of the absolute level of charter school quality depends in part upon quality changes in the traditional public school sector. If, for example, the quality of traditional public schools in Texas is falling over this period due to the expansion of the charter sector or other factors, the catch-up of charter schools may not indicate much if any quality improvement. Alternatively, if traditional public schools improve – either in response to competition from the charter sector or for other reasons – the observed increase in charter school quality would actually understate the improvement in charter school effectiveness. Imberman (2011) highlights the difficulty of identifying the causal effect of competition on traditional schools resulting from charter schools. Therefore, we simply describe changes over time in state average achievement to provide a context for the relative improvement of the charter sector.

During the sample period, the general increase in scores on the National Assessment of Educational Progress (NAEP) suggests a positive change over time in the quality of public education in Texas. The average NAEP score improved from 2000-2011 in fourth and eighth grade mathematics and from 1998-2011 in fourth grade reading (NCES 2015); the average NAEP score remained roughly constant in 8th grade reading during this period.[29] Given the increase over time in the minority enrollment share and the lower average scores of blacks and Hispanics than whites, the improvements in the overall average NAEP scores may well underestimate the gains in school quality. Looking at subgroups, whites, blacks, and Hispanics each improved over this period on all of the NAEP tests including eighth grade reading (NCES 2015).[30] Thus, the relative improvement of charter schools is not driven by a decline in the average quality of traditional public schools, rather our estimates may understate the gains in absolute performance.

2. Entry, Exit, and Improvement

Table 13 disaggregates these overall trends by providing a description of performance changes associated with entry, market (i.e., voluntary) closures, authorizer closures, and improvement.[31] The average increases of 0.12 and 0.09 standard deviations in mathematics and reading, respectively, are

---

[29] NAEP is a national test, often called the "Nation's Report Card," given to representative samples of students in all states. It has reported state performance in math and reading at grades 4 and 8 every two to four years since 1992. Eighth grade reading tests were not available until1998.

[30] Note that schools across the country also tended to improve on these tests over the period, perhaps indicating the impact of federal accountability legislation (No Child Left Behind, or NCLB). Nonetheless, Texas students as a whole and across the racial/ethnic subgroups generally improved more than the national average over this period.

[31] As noted, the tests changed in 2003. Appendix C provides a similar description for just the 2004-2011 period when the TAKS test was used throughout. For this shorter period, the same patterns of charter school improvement hold, although the magnitudes of change are smaller.

attributable to a combination of: (1) improvement in charter schools that persist throughout the period (Panel A); (2) the disproportionate closure of low value-added schools (Panels B and C); and (3) an average value-added of new schools that far exceeds that of the schools that closed (Panel D). Value-added increased by 0.10 standard deviations in math and 0.04 standard deviations in reading for schools that remained open throughout the entire period. The difference between the average value-added of all schools that closed during the period and those that entered equals 0.21 standard deviations in math and 0.24 standard deviations in reading, though schools forced to close by the authorizer actually had somewhat higher average value-added in mathematics than those that closed voluntarily. (The gap in reading was minimal). Notice that the contribution of entrants is amplified by the large number of entrants relative to the number of charter schools continuously open and relative to the number that closed between 2001 and 2011.

The much higher average value-added of entrants compared to exits suggests systematic differences in the quality of charter management organizations that expanded relative to those that contracted. To examine this relationship more closely, we construct a panel that identifies the annual number of schools operated by each charter management organization over the period. We directly estimate the relationship between CMO expansion and quality using regressions of the change in the number of schools operated in year t on the average mathematics and reading value-added of the CMO operated schools in year t-1and year fixed effects. Columns 1 and 2 of Table 14 show a strong, positive relationship between the change in the number of schools and average mathematics and reading value-added of the schools operated in the previous year that is robust to the inclusion of CMO fixed effects. This pattern is consistent with the notion that quality affects demand for a CMO's schools and that CMOs respond in part by expansion or contraction of the number of schools in operation.

The remaining columns explore the possibility of asymmetry in CMO expansion and contraction. Columns 3 and 4 report linear probability models estimating the relationship between CMO average prior year value-added and the probability that the CMO increased the number of schools operated (Columns 3 and 4) and the probability that the CMO decreased the number of schools operated (Columns 5 and 6). Although positive, the coefficients on mathematics and reading value-added in the expansion models of

Columns 3 and 4 are small and insignificant. In contrast, the coefficients in Columns 5 and 6 reveal a larger and significant negative relationship between the prior average value-added and the probability that the number of schools operated by a CMO declined.

The overall pattern is consistent with higher-quality CMOs increasing their market share over time. Importantly, the estimates remain significant even for models that include CMO fixed effects, so they are not just reflecting significant behavioral differences among CMOs.

B. Matching Estimates

We now replicate the previous analysis using the estimates of school value-added generated by the matching models based on Angrist, Pathak, and Walters (2013). Because the pattern of estimates is quite similar we focus on the few salient differences.

1. Performance trends over time

The top pane of Figure 11 shows the changes over time at the 25th, 50th, and 75th percentiles of the charter school mathematics value-added distributions, and similar to the statewide comparisons the figure reveals improvements across the distribution. Unlike the statewide comparisons where gains were concentrated in the lower portion of the distribution and spread throughout the period, however, gains in the matching model appear to be quite similar at the 25th, 50th, and 75th percentiles and concentrated in the first half of the decade. This is certainly consistent with the possibility that charter schools began drawing students from more effective public schools over time, in part because charter schools that closed had drawn students from less-effective public schools. Such improvement in the traditional public school comparison group would produce the observed pattern.

By comparison, the patterns for reading are much more similar across the two models, with improvement concentrated at the 25th percentile of the distribution regardless of the comparison group. Nonetheless, the divergence in timing remains in that the matching model estimates show much smaller charter school gains post-2006, particularly at the 25th percentile. Again, such a pattern is consistent with improvement in the traditional public school comparison group.

2. Entry, Exit, and Improvement

Table 15 disaggregates the contributions of school entry, closures and improvement to the changes over time in the matching method value-added estimates. The patterns in the table are qualitatively and quantitatively similar to those for the statewide comparison for both mathematics and reading. Specifically, schools in operation in both 2001 and 2011 improve somewhat more in mathematics than reading, schools that close voluntarily are drawn from the lower portion of the distribution and much less effective than those that open, and schools closed by the state authorizer are not as ineffective as those that close voluntarily in terms of mathematics value-added. One difference is the very low average reading value-added of schools closed by the authorizer in Table 15; the corresponding value using the state comparison group is much more similar to the average for schools that closed voluntarily. Note that there are six fewer schools in Table 15 due to the absence of matches with non-missing data.

Finally, a comparison of the specifications that estimate the relationship between the changes in number of schools operated and CMO quality reveals a quite similar pattern across the two methods (Table 16). The net change in the number of schools operated is positively related to CMO average VA with a somewhat stronger relationship in reading than in mathematics. A small difference does emerge in the effects on expansion relative to contraction: in comparison to the statewide comparison estimates reported in Table 14, the coefficients in Table 16 reveal less asymmetry.

VIII. EXPLORATORY ANALYSIS OF THE SOURCES OF CHARTER SCHOOL IMPROVEMENT

Recent evidence on the determinants of charter school quality reported in Furgeson et al. (2012), Angrist, Pathak, and Walters (2013), and Dobbie and Fryer (2013) highlights the particularly strong performance of charter schools that set high expectations, require uniforms, or more broadly adopt a No Excuses philosophy. In this section we investigate the relationship between estimates of school quality and some of these same factors in an exploratory analysis of the factors that underlie the dynamic changes we observe. This work builds on the prior analyses by considering a much broader set of schools and highlighting the sensitivity of the estimates to the inclusion of information on student selection and mobility.

Considerable debate surrounds conjectures that selection into and out of charter schools contributes to their performance, and prior evidence on Texas reveals that mobility adversely affects all students, movers and non-movers alike.[32] It is possible, of course, that increasingly positive selection among successive cohorts may affect achievement directly or indirectly through the creation of a more positive classroom environment for all students. However, existing empirical evidence is controversial on the importance and magnitudes of such effects. We control directly for prior achievement and behavior of all students including new entrants. The average prior performance of new entrants relative to their traditional public school classmates conditional on these controls provides some information on peer composition. Nonetheless, the absence of a compelling source of exogenous variation precludes the identification of causal effects of such policies or peer composition, leading us to focus on the descriptive question of the extent to which these student variables can account for the higher performance of schools adhering to a No Excuses philosophy.

The belief that students are inputs into education production in addition to being consumers of its output guides the model of schooling demand in the seminal work by Epple and Romano (1998). It has

[32]Nichols-Barrer et al. (2014) consider the conjecture that student attrition from KIPP schools might explain their success but reject it.

been reinforced by extensive work on peer effects in schools.[33] Informal conversations with CMO executives indicate that many share this belief. These executives, however, tend to emphasize behavior rather than achievement. The No Excuses philosophy encapsulates this theory, often featuring a number of rules or policies including strict discipline, contracts that require parental commitment, and uniforms aimed at creating a positive environment for learning.[34] These rules may contribute to a positive environment both through their direct effects on behavior and through their influence on enrollment and re-enrollment decisions.

We begin with a description of trends over time in the share of schools that adhere to a NO Excuses philosophy to uncover whether the change is consistent with the expansion of the No Excuses model contributing to the improvement in charter sector quality; trends over time in mobility and selection are also presented. Next we report estimates of the relationship between charter school mathematics and reading value-added and adherence to a No Excuses philosophy for a series of specifications that progressively add controls for mobility and selection. Finally, we substitute indicators for the individual components used to classify schools as No Excuses in order to learn more about the underlying sources of any association with value-added. Importantly, the designation of a CMO as adhering to a No Excuses philosophy is not straightforward, as many that appear to operate with rules and practices that correspond to the No Excuses philosophy do not designate themselves in this way. Appendix A describes the extensive information and decision-rules that we use to determine whether a CMO should be classified as following a No Excuses philosophy.

A. Trends over time

Figure 12 shows that, by our measures, the share of students attending charter schools classified as adhering to a No Excuses philosophy increases from roughly 18 to 38 percent between 2001 and 2011. This substantial increase corresponds with the improvement of the charter sector.

[33] See the review in Sacerdote (2011).
[34] See Thernstrom and Thernstrom (2003), Mathews (2009).

The next two figures reveal trends over time that are also consistent with selection and mobility accounting for a portion of the observed improvement in the charter sector, and potentially, the observed association between value-added and a No Excuses philosophy. Figure 13 traces the proportion of charter and traditional public school students that are new to their school. (For this, we restrict the sample to students attending a grade in schools where the prior grade was offered; i.e., the sample excludes students in brand new schools or the first grade offered in a school). Remarkably, the annual share of new students exceeded, on average, 50 percent in charter schools until 2006, reflecting both the enrollment increases experienced by many schools in their early years of operation and frequent movement in and out of charter schools. The percentage of new students, however, declined by almost thirty percentage points between 2001 and 2011; even so, the level remained twice that of the traditional public schools.

To see the changes in composition of the students in charter schools, Figure 14 plots the mean differences in math and reading achievement and the probability of committing a disciplinary infraction between traditional public school students who transition to a charter school in the subsequent year and their schoolmates who remain in the traditional sector.[35] The high rate of charter school mobility shown previously, however, also means that the characteristics of new entrants may not accurately capture the overall degree of selection relevant for on-going school operations. Therefore, while the top panel compares all charter school entrants to schoolmates who remain in the traditional public sector, the bottom panel compares only charter entrants who remain in the charter school into the second year with the same set of schoolmates.

Between 2001 and 2004, the entering achievement and behavior characteristics of charter-school students largely did not improve relative to schoolmates who remained in the traditional public sector, but this picture changed markedly in more recent years for both all entrants and those who remained into their

---

[35]Importantly, all comparisons of achievement and behavior apply to those during the year prior to charter school entry and thus rule out any influences of the charter school. Moreover, disciplinary infraction comparisons within a traditional public school at a point in time hold constant infraction policies and procedures and isolate differences in behavior. For these measures, we first compute the differences between each charter school entrant and her schoolmates who remain in the traditional public sector and then average over the sample of entrants.

second year at the charter. The average difference in mathematics achievement between students who entered a charter school and schoolmates who remained in the traditional sector was -0.23 standard deviations in 2001, fell to -0.30 in 2004, and then rose to -0.05 in 2011; the corresponding differences for reading were -0.20 standard deviations, -0.21, and 0.03, and the corresponding differences in the probability of a disciplinary infraction were 0.06, 0.16, and 0.05. In sum, student selection into charter schools based on achievement moved from being negative in 2001 to roughly neutral in 2011, while selection changed little in terms of behavior.

Entrants who remained in their charter schools into the second year following the transition were less negatively selected in 2001 and generally more positively selected in 2011 than new entrants as a whole, indicating adverse selection out of charters. By comparison, the traditional public school students who remained in their school into the second year were quite similar to those who remained in the traditional sector but switched schools prior to the second year.

B. Value-added, No Excuses, and Selection

For policy, a pressing question is the extent to which this student selection accounts for the higher performance of charter schools that adhere to a No Excuses philosophy or other operational characteristics associated with superior outcomes. In order to understand better the interrelationships among mobility, selection, and adherence to a No Excuses philosophy, we estimate a series of models that regress mathematics or reading value-added on various combinations of these variables.[36] Once again we begin

[36]For this analysis, the selection at the time of entry and reenrollment variables are computed as follows: first, each charter school entrant is assigned the difference between their prior achievement (or receipt of a disciplinary infraction) and the average among their traditional public school peers that remain in the traditional public sector. Next, these differences are averaged over all students that enter each school. The reenrollment selection variables are computed similarly with the exception that the differences are averaged over only those students who remain in the same charter into their second year. For students who enter a charter school in year t, the degree of selection upon entry is related to value-added in year t, while the degree of selection at the time of reenrollment for the second year is related to value-added in year t+1. Standard errors are clustered at the school level; clustering at the CMO level has little effect on the standard errors.

with specifications that use the statewide estimates of value-added and then turn to those that based on local market matching models.

The No Excuses coefficients in both reading and mathematics are sensitive to the inclusion of the student composition variables, though they remain highly significant in all specifications (Table 17). In the mathematics models, the coefficient declines from 0.164 to 0.097 following the inclusion of the student mobility variable and to 0.092 in the specification that also adds the selection on achievement variables. By contrast, including the disciplinary infractions variables does not alter the No Excuses advantage, suggesting that selection on this dimension of student composition does not drive the changing performance of charter schools.

Estimates for reading follow a very similar pattern, declining from 0.091 to 0.028 following the inclusion of the mobility and selection variables. In this case the No Excuses coefficient is smaller and more sensitive to the additional inclusion of the selection on achievement variables.

A comparison between Tables 17 and 18 also reveals a very similar pattern of No Excuses coefficients across the two alternative methods for the estimation of value-added. In the case of mathematics, the No Excuses coefficient from regressions that use the value-added estimates produced by the matching method declines from 0.185 for a specification that does not control for mobility or selection to 0.129 following the inclusion of those controls, while in corresponding specifications for reading the No Excuses coefficient falls from 0.113 to 0.066.

Given the absence of a compelling source of exogenous variation and possibility that these variables capture unobserved student and school differences, the coefficients for the selection and mobility variables do not warrant a causal interpretation. Nonetheless, the insensitivity of the No Excuses coefficient to the inclusion of any of the selection variables in the mathematics specifications suggests that selection accounts for little of the No Excuses effect; more positive selection may contribute to charter school improvement in terms of mathematics value-added, but it does not appear to account for the higher performance of No Excuses schools. In reading the findings are less clear cut and more consistent with selection accounting for some of the association between school quality and adherence to a No Excuses philosophy. Again this is consistent with the notion that schools play a more important role in the determination of achievement in mathematics as opposed to reading.

The large and significant mobility estimates and the sensitivity of the No Excuses coefficients to the inclusion of mobility in combination with the dramatic decline in the average share of students who are new to the school suggest an important role for mobility in the improvement of the charter sector. One approach to quantifying that contribution is to use the causal estimate of mobility externalities from Hanushek, Kain, and Rivkin (2004), also based on Texas data, to estimate the contribution of mobility to the increase in charter school mathematics value-added.[37] That estimate suggests that the approximately 20 percentage point decline in the charter-traditional public school differential in the share of students that are new to the schools contributes roughly 0.04 standard deviations to the improvement of charter school math performance between 2001 and 2011.[38] Thus the greater sector stability *per se* accounts for over one-third of the decrease in the average mathematics value-added gap between charter and traditional public schools. Note, this is an estimate of the externality of high student mobility as the value-added regressions account for the direct effects of moving on individual movers.

The final component of this analysis examines the relationship between value-added and the underlying components of the No Excuses classification using the value-added estimates from the state comparisons (Table 19) and from the matching models (Table 20). A comparison of the two tables reveals many similarities but also some differences. In terms of similarities, the inclusion of the selection and mobility variables reduces the coefficients on the No Excuses components virtually across the board, though the magnitude of the reduction tends to be larger for mathematics and much larger for the state comparison method. Consequently, both the uniform and dress code indicators remain significant for reading using the

[37]Note that estimates of the impact of mobility externalities are not available for reading
[38]Hanushek, Kain, and Rivkin (2004) find that the added disruption of high mobility creates an externality. That analysis is based on value-added models of achievement in Texas that include student, school-by-year, and school-by-grade fixed effects to account for confounding factors including perceived school quality and neighborhood shocks. A ten percentage point higher level of mobility reduces mathematics achievement by approximately 0.2 standard deviations in Texas public schools (independent of any impact on the individuals who move).

matching but not using the state comparison method; the uniform indicator does remain significant regardless of method for the mathematics specifications.

The finding that value-added has the strongest relationship with the uniform requirement is consistent with the prior research, as is the positive relationship between value-added and high expectations. Note, however, that the high expectations coefficient is significant at only the 10 percent level in some specifications and not even at that level in others. Because a uniform requirement is straightforward to measure in comparison to classifications that require more discretion, the absence of measurement error may elevate the magnitude and significance of the uniform coefficient in comparison to others. This merits additional investigation, but the consistent pattern across cities and strength of the relationship in specifications that account for mobility and selection suggests that the enforcement of a uniform requirement is associated with higher school value-added.

C. Other Contributing Factors

Classification as a No Excuses school is, of course, not the sole important dimension of school operations, and there are certainly others, most notably the quality of leadership and instruction, that vary among schools regardless of their philosophy. In fact, informal conversations with several executives employed by some of the largest CMOs operating in Texas, including several from No Excuses CMOs, reveal a strong emphasis on the hiring and development of effective school leaders. Some CMOs devote substantial resources to the identification and training of school leaders including year-long apprenticeships. These preparation programs differ considerably from the traditional public school job ladder of teacher to assistant principal to principal combined with some formal education in leadership. Other CMOs bemoaned the inability to afford such programs. Importantly, this commitment to leadership did not seem to depend on the degree of authority granted over personnel or programmatic decisions. Impediments to the measurement of leadership performance complicate the identification of its contribution to charter school quality and improvement, and this is a prime area for further investigation.[39]

[39] See Branch, Hanushek, and Rivkin (2012)and Laing et al (2015) on both the potential importance of principals and the difficulty of measuring differences among principals. Bloom et al. (2014) also point to the importance of management in schools, relying on surveys of specific management practices. England has introduced Academy Schools which call for conversion of traditional public schools into institutions very similar to charter schools, and this has led to positive but heterogeneous impacts on student performance. When surveyed, a majority of the Academy Schools indicated that change in leadership was the most important element of their conversion; see Eyles and Machin (2014).

IX. CONCLUSION

This paper uses administrative data from Texas to trace the evolution of charter school quality and to estimate the effects of ratings assigned by the statewide school accountability rating system. The main highlights of this study are that the incentive effects of school accountability are concentrated at lower rating levels where statutory consequences are triggered, and that the evaluation of charter schooling laws strictly on the basis of contemporaneous achievement differences misses the dynamic improvements within the sector.

The important policy take away from the school accountability analysis is that ratings effects are concentrated at the low end of the school quality distribution where state interventions are triggered. Rating differentiation at the margin between schools at the upper end appears not to be informative nor do the incentives appear strong enough to motivate improvement. Whether this is due to the low average quality of the schools themselves or because this is the only margin with direct consequences can not be separated. The natural policy prescription is a simplification of the four-level rating system to a two tiered system with ratings equivalent to Pass and Fail. This in fact is what was implemented in Texas in 2014. Along with a transition to a new assessment regime, schools are now rated acceptable or needs improvement. Evaluation of this new system will be informative to see if the differential response between the two new ratings is similar to the old acceptable-unacceptable boundary, or if the new effect represents an average of the exemplary, recognized and acceptable effects.

The charter school component of this dissertation provides guidance to two aspects of charter school policy. The first is motivated by the notion that dynamic changes within the charter sector form part of the ex-ante motivation for chartering. Any evaluation based strictly on the quality of an immature charter school or charter sector at a single point in time is likely to miss this critical component of the effectiveness of a charter school policy. Second, lottery based estimates of school quality are important for learning about the effectiveness of particular overprescribed schools. Importantly however, a component of charter policy inherently relies on lifting the constraints placed on school administrators. Any attempt on the part of policy makers to favor certain pedagogical or organizational approaches that are found to be effective by lottery studies potentially restricts the potential for future innovation. It is important to reiterate that lottery studies provide vital information about effective practices, but this information is more relevant for prospective or current school operators, rather than for current charter school policy makers. Over time, however, it may become apparent that policy makers should consider the explicit incorporation of practices that are found to be effective at charter schools into the traditional public school sector.

I now summarize the findings and implication of the school accountability and charter school analyses in further detail in turn.

A. School Ratings

The school rating system in Texas is not unique in its reduction of a variety of continuous performance metrics into a single discrete measure of quality. By unpacking the rating assignment system, I exploit the cutoffs between rating levels to find a series of quasi-experiments that yield plausibly random rating assignments which can be used to estimate the effect of being assigned each rating. These 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 families and school administrators. In this paper I supply evidence of 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 in ways to improve performance? This study adds to the evidence that accountability ratings can induce improvements for low performing schools, while ratings at higher achieving school provide little incentive for schools to improve.

Schools' responses to rating shocks may be viewed in the context of prior literature on organizational responses to external stresses. The issue of organizational responses to external stresses in the school context is put forward by Smylie (2009). For many schools, testing and the resulting ratings make up a fundamental driver of school operations. In this context an adverse rating outcome could be viewed as an external stress. While I find that rating "stresses" result in positive outcomes, it is possible that these outcomes are achieved through "maladaptive tendencies." In this situation, this could include educational triage, changes antithetical to longer term improvements, and increased centralization. The degree to which these different responses occur offers fertile ground for future work.

The generally negative, though small, effect on reenrollment of a lower rating confirms that the information content of a rating is only of little marginal value to families. On the other hand, the finding that low rated schools show improved performance and negative point estimates on reenrollment suggests that perhaps the indirect reputation effects and the direct intervention effects may be offsetting. Nevertheless, extending this analysis to the response of families with no prior exposure to a particular school may help shed light on the relative strengths of these two mechanisms.

B. Charter Schools

Charter quality is measured by value-added to mathematics and reading. The results based on quality measures generated by a flexible value-added specification that controls for prior achievement and discipline and those generated by a matching procedure based on the specification used in Angrist, Pathak, and Walters (2013) lead to the same finding: charter-school mathematics and reading value-added increased substantially relative to those for traditional public schools. This improvement is notable because there is evidence that traditional public schools were also improving on average over this time period.

The main charter school results also support the belief that market forces are generating dynamic improvements in the charter sector. First, consistent with existing evidence, schools that close are drawn disproportionately from the bottom tail of the charter school distribution. Second, schools that open during the study period far outperform those that CMOs decide to close, with average value-added for new charters roughly equal to the average for existing charters. Finally, average value-added increases for charter schools that remain open throughout the period. Together these changes raise the mean and reduce the variance of school value-added relative to traditional public schools.

Examining the potential sources of these improvements, we find evidence that an increasing share of schools that adhere to a No Excuses philosophy and a reduction in student mobility as the sector matures contribute to the charter sector's improvement, though substantial portions of the improvement remains unexplained by these factors. Looking more closely at the first, the pattern of estimates suggests that student selection and mobility account for a portion of the No Excuses premium. Nevertheless, the No Excuses coefficient remains highly significant in all specifications, and additional controls for selection on behavior have little or no effect on the magnitude of the estimates. As a whole the findings are consistent with the belief that the schools adhering to a No Excuses philosophy are more effective on average, though the precise mechanisms are not identified. Additional investigation finds that a uniform requirement has the strongest relationship with value-added among the limited number of variables used to determine No Excuses status, but this may be driven in part by the precision with which this factor is measured.

The substantial decline in student mobility and the increasingly positive selection of new entrants to the charter sector provide valuable information on sector dynamics and the importance of patience in an evaluation of a large-scale educational reform, particularly one that relies on parental choices and market forces. The relaxation of constraints on school management induced many with little prior experience to enter into public education, and the large variation in school quality observed during the early years are consistent with growing pains associated with a new market. These factors likely contribute to the high student mobility and the unwillingness of many students making adequate progress in the traditional public schools to consider a switch to a charter

Over time, many low-performing schools closed and the average effectiveness of new market entrants and schools remaining open throughout the decade rose. As might be expected, students and families appear to respond favorably to these improvements. Over time selection into the charter sector became less negative, as the prior achievement of new entrants rose relative to their classmates who chose to remain in the traditional public school sector. Thus the families of higher-achieving students appear to have elevated their opinion of a charter schools as a viable alternative. Importantly, these responses amplified the improvements in the sector by raising the quality of the classroom environment.

The juxtaposition of these dynamic changes with cross-sectional comparisons of sector differences highlights the value of a focus on the trajectory of school quality as opposed to effectiveness at a point in time in the evaluation of a major educational reform. Much more can be learned about the behaviors of both families and education providers and the aspects of school operations that contributed to the improvement. Although the identification of the contributions of specific school factors including the quality of teachers, principals, and CMO executives may be difficult, this is a prime area for additional research.

REFERENCES

Abdulkadiroğlu, Atila, Joshua D. Angrist, Susan M. Dynarski, Thomas J. Kane, and Parag A. Pathak. 2011. "Accountability and flexibility in public schools: Evidence from Boston's charters and pilots." *Quarterly Journal of Economics* 126, no. 2: 699-748.

Ahn, Thomas, and Jacob Vigdor. 2014. "The Impact of No Child Left Behind's Accountability Sanctions on School Performance: Regression Discontinuity Evidence from North Carolina." *NBER Working Paper 20511*. September.

Angrist, Joshua D., Susan M. Dynarski, Thomas J. Kane, Parag A. Pathak, and Christopher R. Walters. 2012. "Who Benefits from KIPP?" *Journal of Policy Analysis and Management* 31, no. 4 (Fall): 837-860.

Angrist, Joshua D., Parag A. Pathak, and Christopher R. Walters. 2013. "Explaining Charter School Effectiveness." *American Economic Journal: Applied Economics* 5, no. 4 (October): 1-27.

Ball, Stephen J., and 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): 377-400.

Bifulco, Robert, and Helen F. Ladd. 2006. "The impacts of charter schools on student achievement: Evidence from North Carolina." *Education Finance and Policy* 1, no. 1 (Winter): 50-90.

Bloom, Nicholas, Renata Lemos, Raffaella Sadun, Daniela Scur, and John Van Reenen. 2014. "The New Empirical Economics of Management." Centre for Economic Performance Occasional Paper 41. London: London School of Economics and Political Science (April).

Booker, Kevin, Scott M. Gilpatric, Timothy Gronberg, and Dennis Jansen. 2007. "The impact of charter school attendance on student performance." *Journal of Public Economics* 91, no. 5-6: 849-876.

Branch, Gregory F., Eric A. Hanushek, and Steven G. Rivkin. 2012. "Estimating the Effect of Leaders on Public Sector Productivity: The Case of School Principals." NBER Working Paper W17803. Cambridge, MA: National Bureau of Economic Research (January).

Chetty, Raj, John N. Friedman, and Jonah Rockoff. 2014a. "Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates." *American Economic Revew* 104, no. 9 (September): 2593–2632.

Chetty, Raj, John N. Friedman, and Jonah Rockoff. 2014b. "Prior Test Scores Do Not Provide Valid Placebo Tests of Teacher Switching Research Designs." (September).

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

CREDO. 2009. *Multiple choice: Charter school performance in 16 states*. Stanford, CA: Center for Research on Education Outcomes, Stanford University.

CREDO. 2013. *National charter school study 2013*. Stanford, CA: Center for Research on Education Outcomes, Stanford University.

Dee, Thomas and Brian Jacob. 2009. "The impact of NCLB on student achievement." *NBER Working Paper No. 15531*.

Deming, David J. 2014. "Using School Choice Lotteries to Test Measures of School Effectiveness." *American Economic Review* 104, no. 5 (May): 406-11.

Dobbie, Will, and Roland G. Fryer. 2011. "Are High-Quality Schools Enough to Increase Achievement among the Poor? Evidence from the Harlem Children's Zone." *American Economic Journal: Applied Economics* 3, no. 3 (July): 158-87.

Dobbie, Will, and Roland G. Fryer, Jr. 2013. "Getting beneath the veil of effective schools: Evidence from New York City." *American Economic Journal: Applied Economics* 5, no. 4 (October): 28-60.

Epple, Dennis, and Richard E. Romano. 1998. "Competition between private and public schools, vouchers, and peer-group effects." *American Economic Review* 88, no. 1 (March): 33-62.

Eyles, Andrew, and Stephen Machin. 2014 of Conference. "The Introduction of Academy Schools to England's Education." Paper presented at *CESifo Area Conference on Economics of Education*, September 12-13, at Munich.

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): 591-604.

Fortson, Kenneth, Natalya Verbitsky-Savitz, Emma Kopa, and Philip Gleason. 2012. *Using an Experimental Evaluation of Charter Schools to Test Whether Nonexperimental Comparison Group Methods Can Replicate Experimental Impact Estimates*. Washington, DC: Institute for Education Sciences (April).

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.

Furgeson, Joshua, Brian Gill, Joshua Haimson, Alexandra Killewald, Moira McCullough, Ira Nichols-Barrer, Bing-ru Teh, Natalya Verbitsky-Savitz, Melissa Bowen, Allison Demeritt, Paul Hill, and Robin Lake. 2012. *Charter-School Management Organizations: Diverse Strategies and Diverse Student Impacts*. Princeton, NJ: Mathematica Policy Research.

Gleason, Philip, Melissa Clark, Christina Clark Tuttle, and Emily Dwoyer. 2010. *The Evaluation of Charter School Impacts: Final Report*. Washington, DC: National Center for Education Evaluation and Regional Assistance (June).

Guarino, Cassandra M., Mark D. Reckase, and Jeffrey M. Wooldridge. 2012. "Can Value-Added Measures of Teacher Performance be Trusted?" Education Policy Center Working Paper #18. East Lansing, MI: Michigan State University (May).

Guarino, Cassandra M., Mark D. Reckase, and Jeffrey M. Wooldridge. 2015. "Can Value-Added Measures of Teacher Performance Be Trusted?" *Education Finance and Policy* 10, no. 1: 117-156.

Hanushek, Eric A., John F. Kain, and Steve G. Rivkin. 2004. "Disruption versus Tiebout improvement: The costs and benefits of switching schools." *Journal of Public Economics* Vol 88/9-10: 1721-1746.

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

Hanushek, Eric A., John F. Kain, Steve G. Rivkin, and Gregory F. Branch. 2007. "Charter school quality and parental decision making with school choice." *Journal of Public Economics* 91, no. 5-6 (June): 823-848.

Hanushek, Eric A. and Margaret F. Raymond. 2005. "Does School Accountability Lead To Improved Student Performance?," *Journal of Policy Analysis and Management*, v24 (2,Spring): 297-327.

Hanushek, Eric A., Jens Ruhose, and Ludger Woessmann. 2015. "Economic Gains for U.S. States from Educational Reform." NBER Working Paper Cambridge, MA: National Bureau of Economic Research (November).

Hastings, J., and Jeffrey Weinstein. 2008. "Information, school choice and academic achievement: evidence from two experiments", *Quarterly Journal of Economics* 123, Issue 4, 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.

Hoxby, Caroline M., Sonali Murarka, and Jenny Kang. 2009. *How New York City's Charter Schools Affect Achivement*. Cambridge, MA: New York City Charter Schools Evaluation Project (September).

Imbens, G., and Karthik Kalyanaraman. 2012. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator," *Review of Economic Studies*, **79**(3), 933–959.

Imberman, Scott A. 2011. "The effect of charter schools on achievement and behavior of public school students." *Journal of Public Economics* 95, no. 7-8 (August): 850-863.

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

Jacob, Brian, Lars Lefgren, and David Sims. 2010. "The Persistence of Teacher-Induced Learning Gains." *Journal of Human Resource.s* 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.

Mathews, Jay. 2009. *Work hard. Be Nice.: How two inspired teachers created the most promising schools in America*. Chapel Hill, NC: Algonquin Books.

McCaffrey, Daniel F., Tim R. Sass, J. R. Lockwood, and Kata Mihaly. 2009. "The Intertemporal Variability of Teacher Effect Estimates." *Education Finance and Policy* 4, no. 4 (Fall): 572-606.

Meghir, Costas, and 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: 1-87.

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

Neal, Derek, and Diane Whitmore Schanzenbach. 2010. "Left Behind by Design: Proficiency Counts and Test-Based Accountability," *Review of Economics and Statistics*, 92:2, 263-283.

Nichols, Austin. 2011. rd 2.0: Revised Stata module for regression discontinuity estimation. http://ideas.repec.org/c/boc/bocode/s456888.html

Nichols-Barrer, Ira, Brian P. Gill, Philip Gleason, and Christin Clark Tuttle. 2014. "Does student attrition explain KIPP's success?" *Education Next* 14, no. 4 (Fall).

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.

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

Rothstein, Jesse. 2014. "Revisiting the impacts of teachers " (mimeo) Department of Economics: University of California, Berkelety (October).

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.

Sacerdote, Bruce. 2011. "Peer effects in education: How might they work, how big are they and how much do we know thus far?" In *Handbook of the Economics of Education, Vol. 3*, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann. Amsterdam: North Holland: 249-277.

Sass, Tim R. 2006. "Charter schools and student achievement in Florida." *Education Finance and Policy* 1, no. 1 (Winter): 91-122.

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.

Smylie, M.A., 2009. Continuous school improvement. Corwin Press.

Texas Education Agency, "2007 Accountability Manual," Department of Assessment, Accountability, and Data Quality Division of Performance Reporting, May 2007.

Thernstrom, Abigail, and Stephan Thernstrom. 2003. *No excuses: Closing the racial gap in learning*. New York, NY: Simon & Schuster.

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, 91–136.

U.S. Department of Education. 2014. *Digest of Education Statistics, 2013*. Advanced Release ed. Washington, DC: National Center for Education Statistics.

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): 844-856

Wong, Manyee, Thomas D. Cook, and Peter Steiner. 2009. “NCLB: An interim evaluation of its effects on learning using two interrupted time series each with its own non-equivalent comparison series.” NW Institute for Policy Research Working Paper 09-11.

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

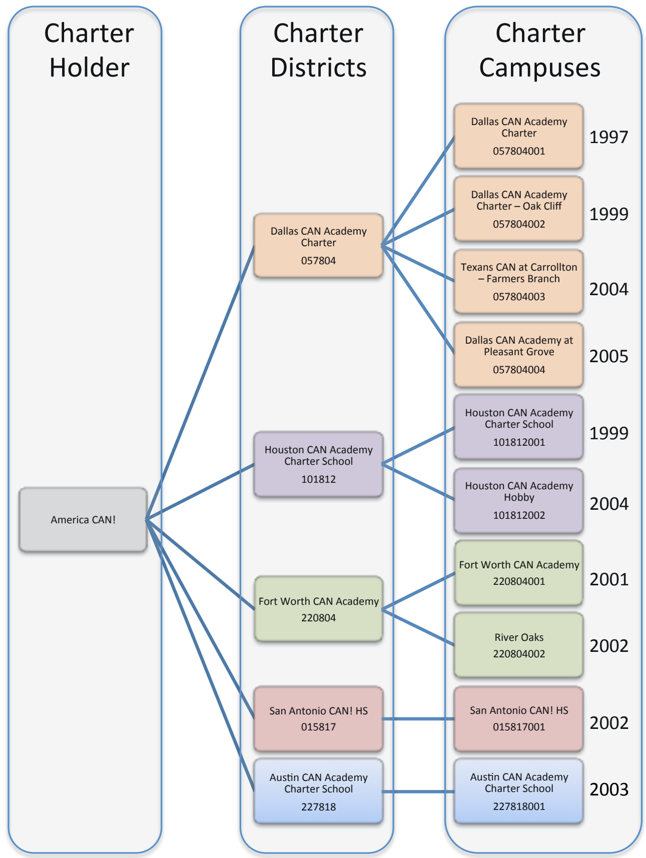


Figure 1: An example of the charter sector organizational structure: the expansion of the America Can! CMO from 1997-2011.

Note: The number in each district and campus block refers to the relevant state ID code.

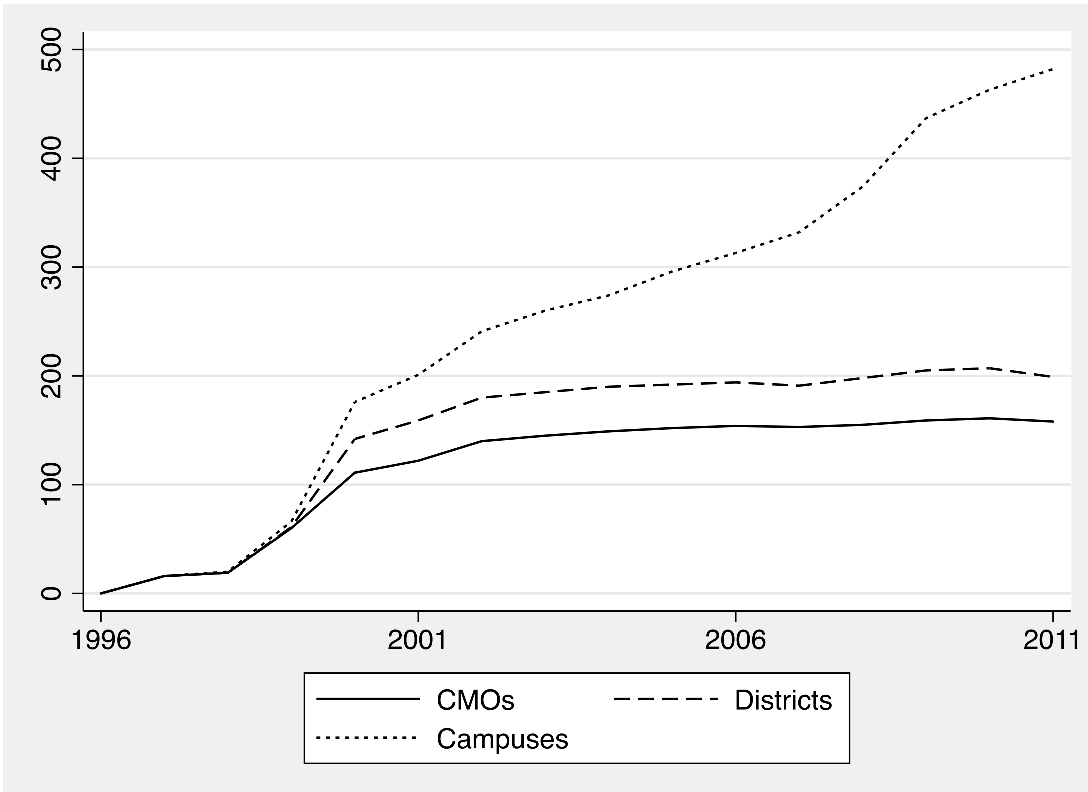


Figure 2: The Growth in Open-Enrollment Charter School, 1995-2011

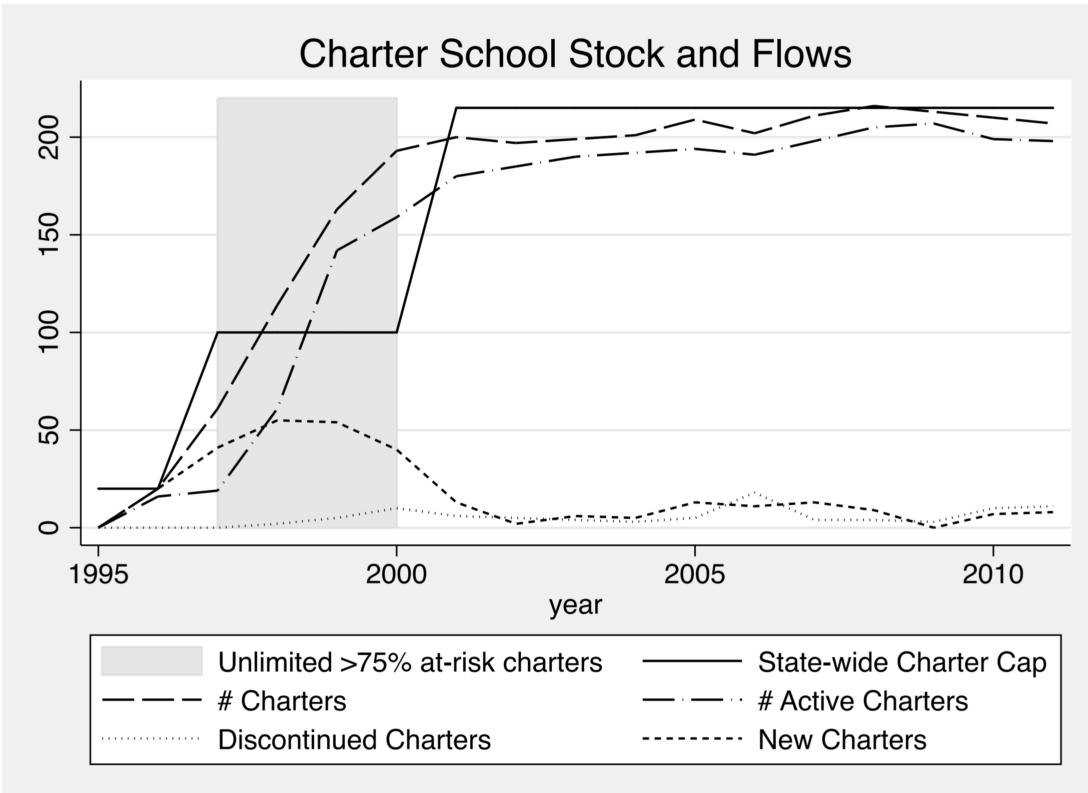


Figure 3. Stock and Flows of State Charters by Type, 1995-2011

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

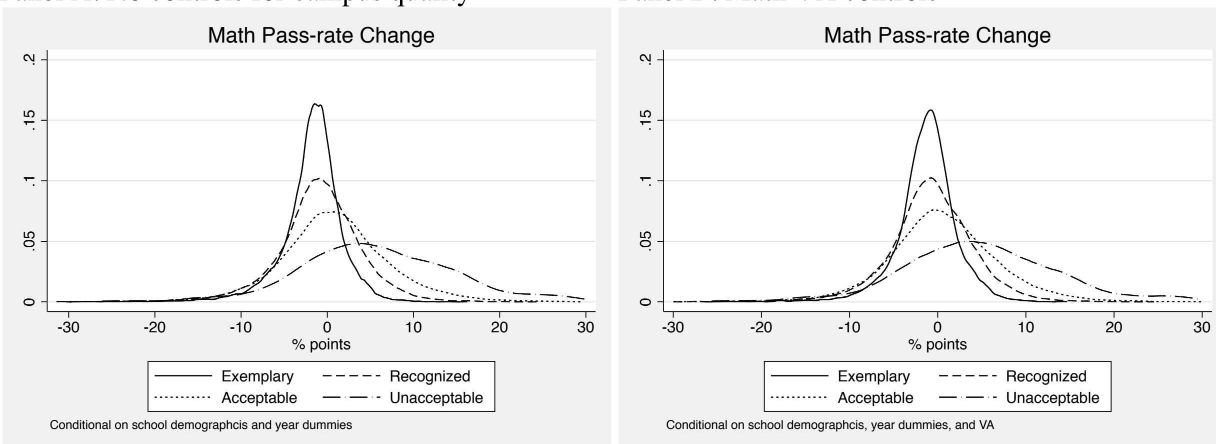


Figure 4. Residual Math Pass-Rate Distributions by lagged school rating.

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

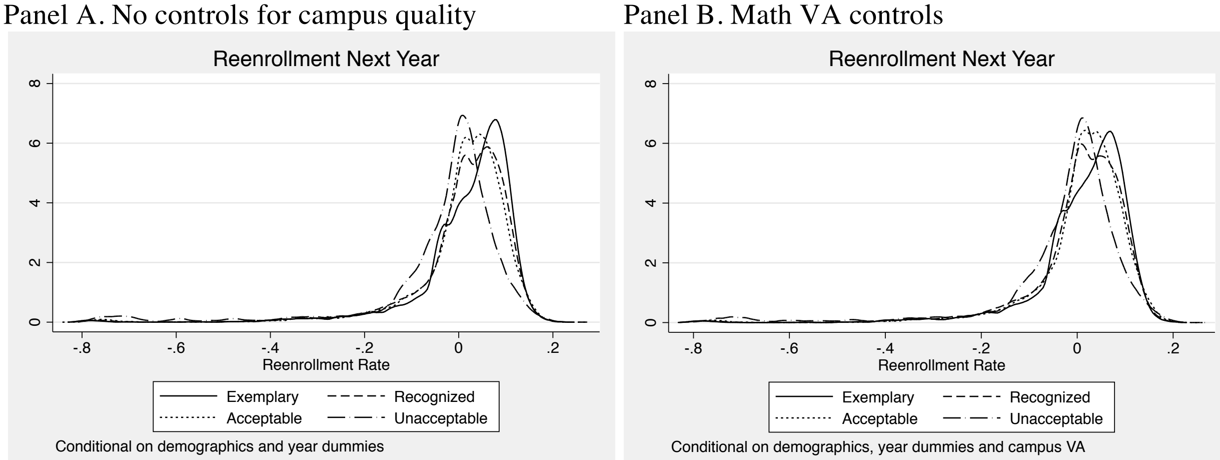
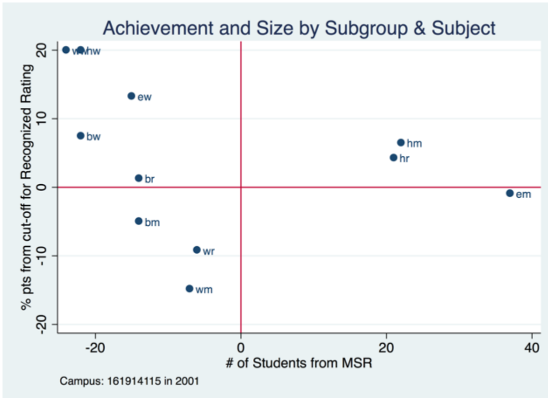


Figure 5. Residual Reenrollment Distributions by lagged school rating.

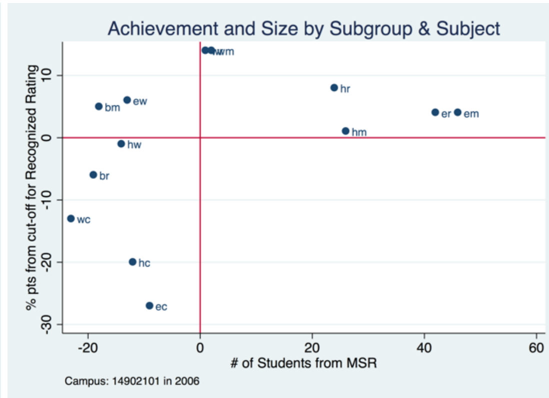
Notes: Figures are based the residuals from regressions of one year ahead reenrollment on controls for student demographics and year indicators. Panel B adds a linear control for campus quality using estimates of campus value added to math achievement.

Pass Rate Discontinuity: Boundary between Quadrants I-IV

Panel A. Treated

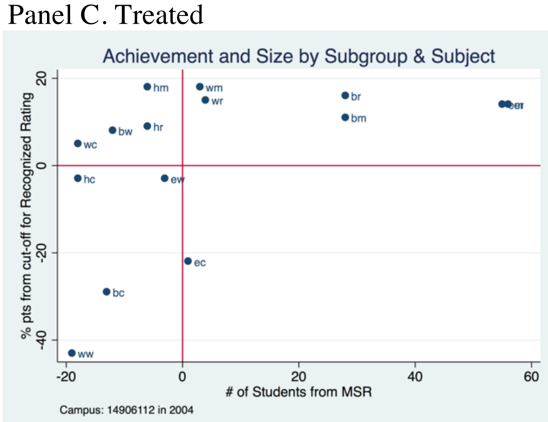


Panel B. Untreated



Group Size Discontinuity: Boundary between Quadrants III-IV

Panel C. Treated



Panel D. Untreated

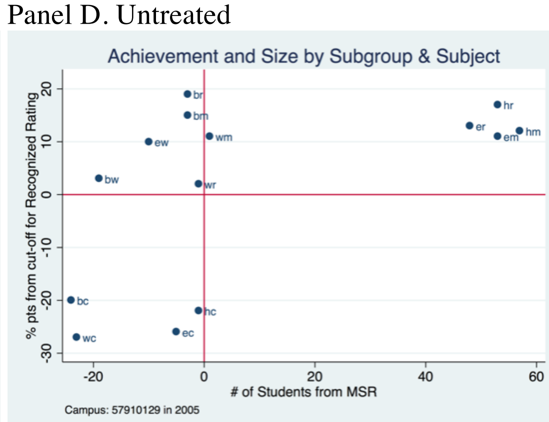
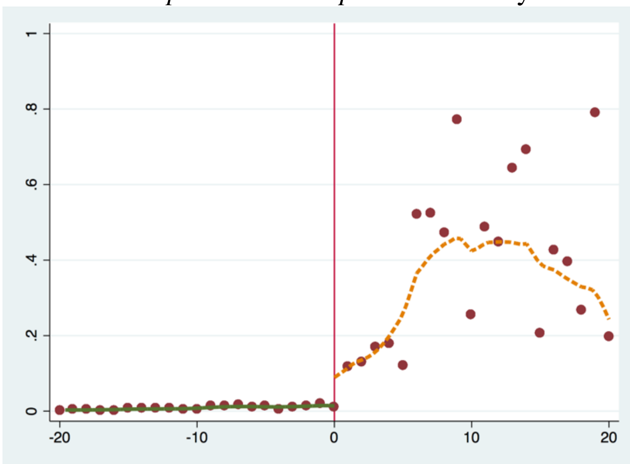


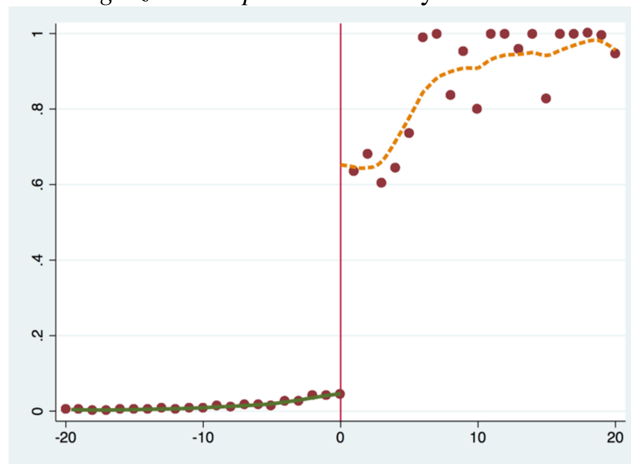
Figure 6. Example campuses by treatment status.

Notes: Each panel plots the size and pass rate of each student group in each subject relative to the pass-rate requirement for the recognized rating and minimum size requirement for evaluation, respectively, for a specific campus in a specific year. Panels (a) and (b) illustrate campuses within a small bandwidth of one percentage point of the pass-rate discontinuity. The campus in panel (a) is is treated because its economically disadvantage students failed to meet the pass rate standard for math while being large enough to count (dot *em*). The campus in panel (b) is is untreated because its Hispanic students failed to meet the pass rate standard for math while being large enough to count (dot *hm*). Panels (c) and (d) illustrate campuses within a bandwidth of one student of the group size discontinuity. The campus in panel (c) is treated because its economically disadvantaged student group was large enough to count while failing to meet the standard for science (dot *ec*). The campus in panel (d) is untreated because its Hispanic student group was large enough to count while failing to meet the standard for science (dot *hc*).

Panel A. *Acceptable-Unacceptable* boundary



Panel B. *Recognized-Acceptable* boundary



Panel C. *Exemplary-Recognized* boundary

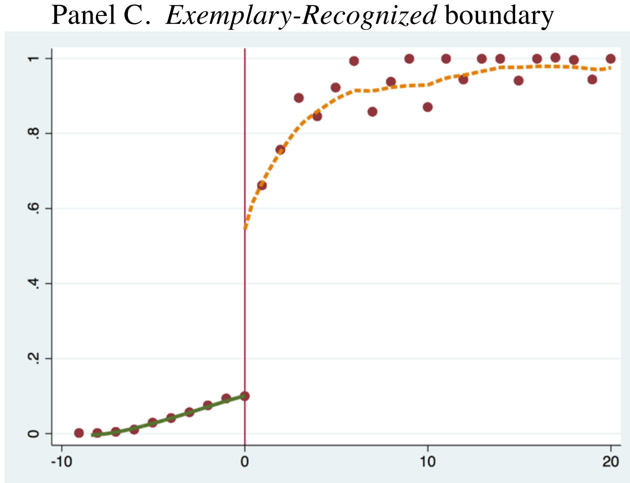
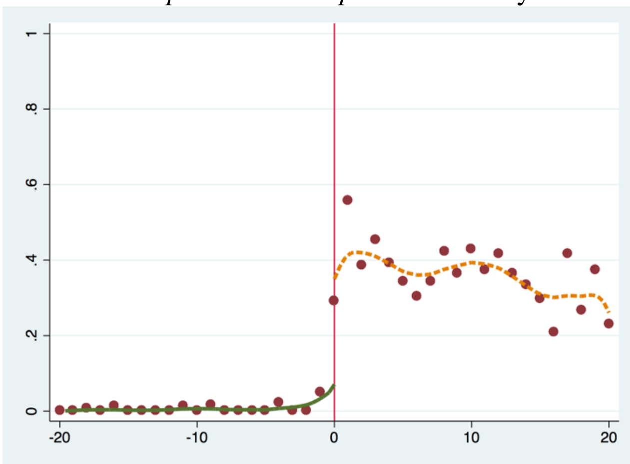


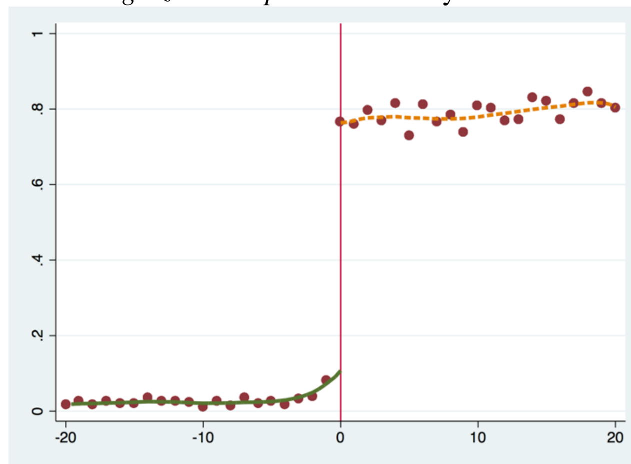
Figure 7. First Stages: Group Pass-Rate Instruments

Notes: Each panel gives the proportion of schools rated below the specified rating boundary by the schools' worst performing large group's percentage point distance from the pass rate requirement for that subject. The horizontal axis is expressed as the number of percentage points *below* the cut-off. In this way, schools to the right of the boundary have a low performing group that is subject to evaluation and are treated with a higher probability of receiving a lower rating.

Panel A. *Acceptable-Unacceptable* boundary



Panel B. *Recognized-Acceptable* boundary



Panel C. *Exceptional-Recognized* boundary

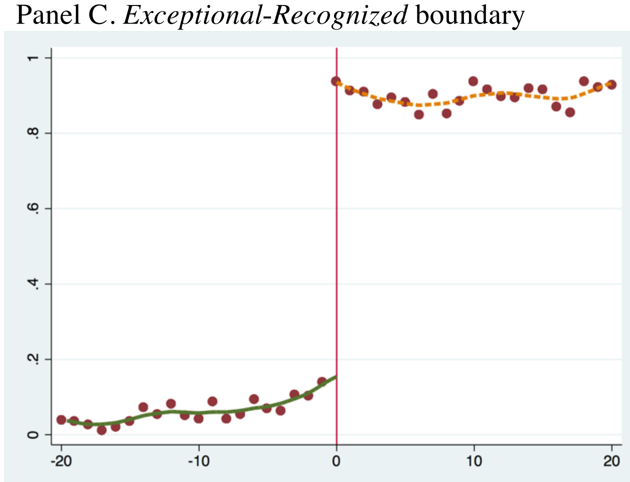
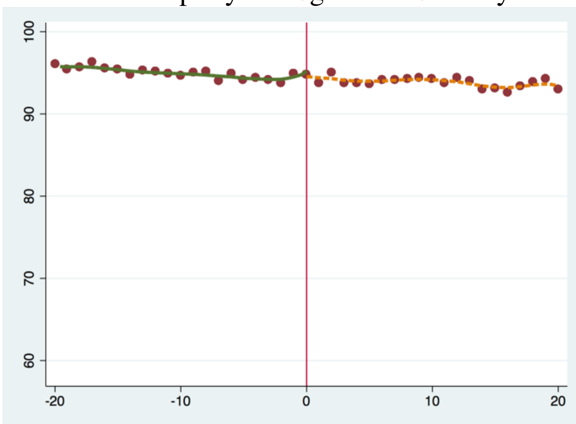


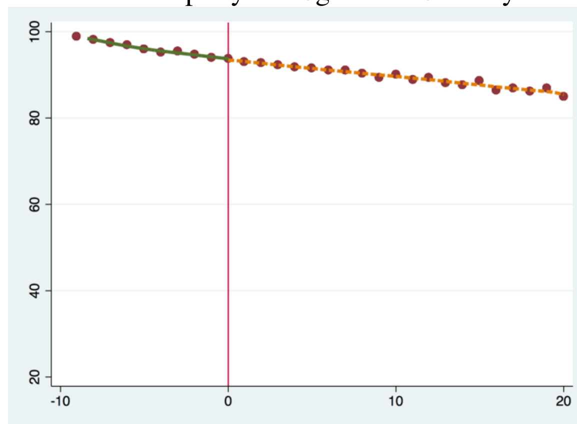
Figure 8. First Stages: Group Size Instruments

Notes: Each panel gives the proportion of schools rated below the specified rating boundary by the schools' largest underperforming group's number of student distance from the minimum size requirement for that group to be evaluated in that subject. The horizontal axis is expressed as the number of students *above* the cut-off. In this way, schools to the right of the boundary have a low performing group that is subject to evaluation and are treated with a higher probability of receiving a lower rating.

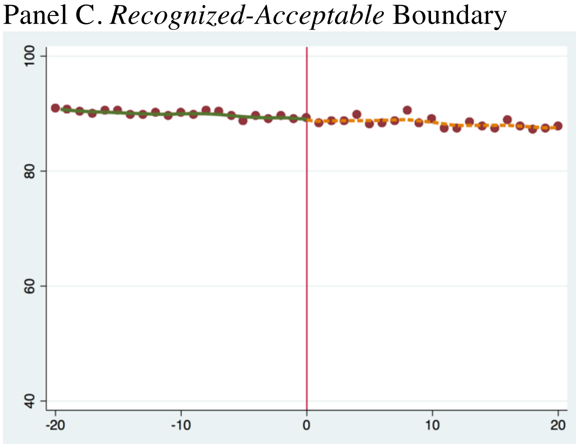
Panel A. Exemplary-Recognized Boundary



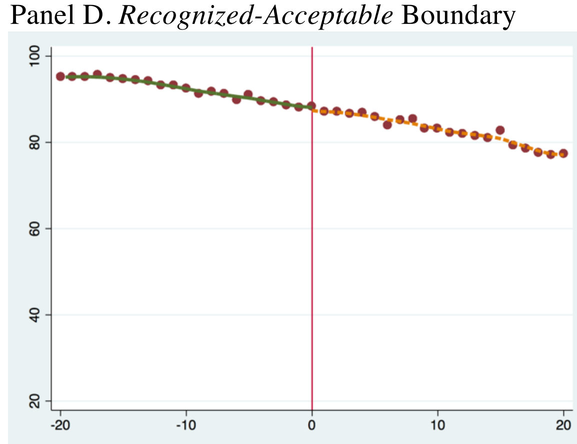
Panel B. Exemplary-Recognized Boundary



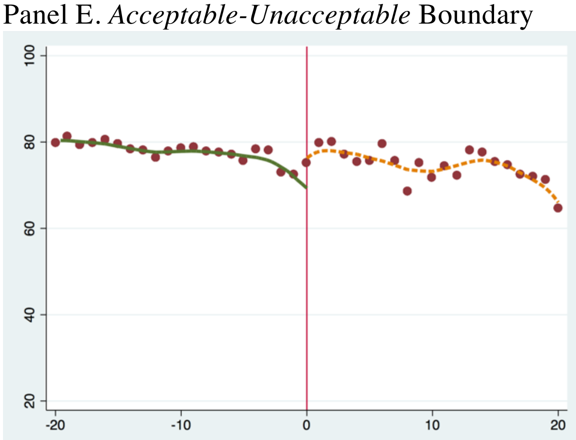
Panel C. *Recognized-Acceptable* Boundary



Panel D. *Recognized-Acceptable* Boundary



Panel E. *Acceptable-Unacceptable* Boundary



Panel F. *Acceptable-Unacceptable* Boundary

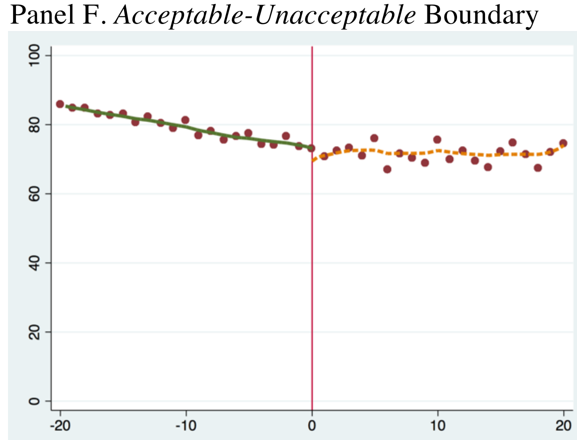
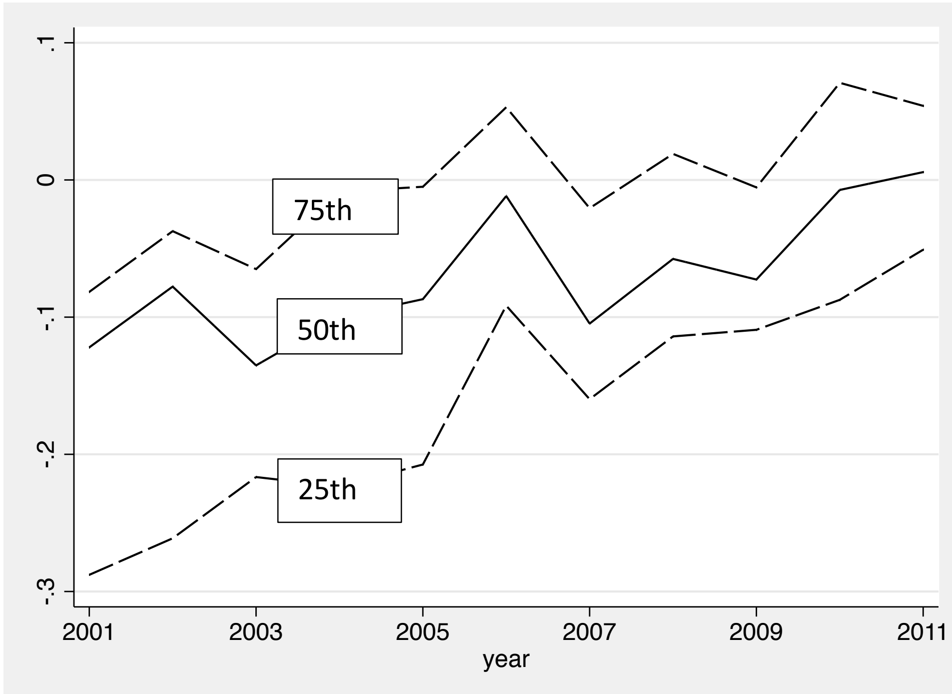


Figure 9. Reduced Form – Math Pass Rate

Notes: In all figures, the vertical axis is the math pass rate the following year. Left hand figures give the reduced form effects on math using the group size instrument with the horizontal measuring the number of students *above* the boundary. Figures on the right give use the pass rate instrument with the horizontal measuring the number of percentage points *below* the standard.

Panel A: Mathematics



Panel B: Reading

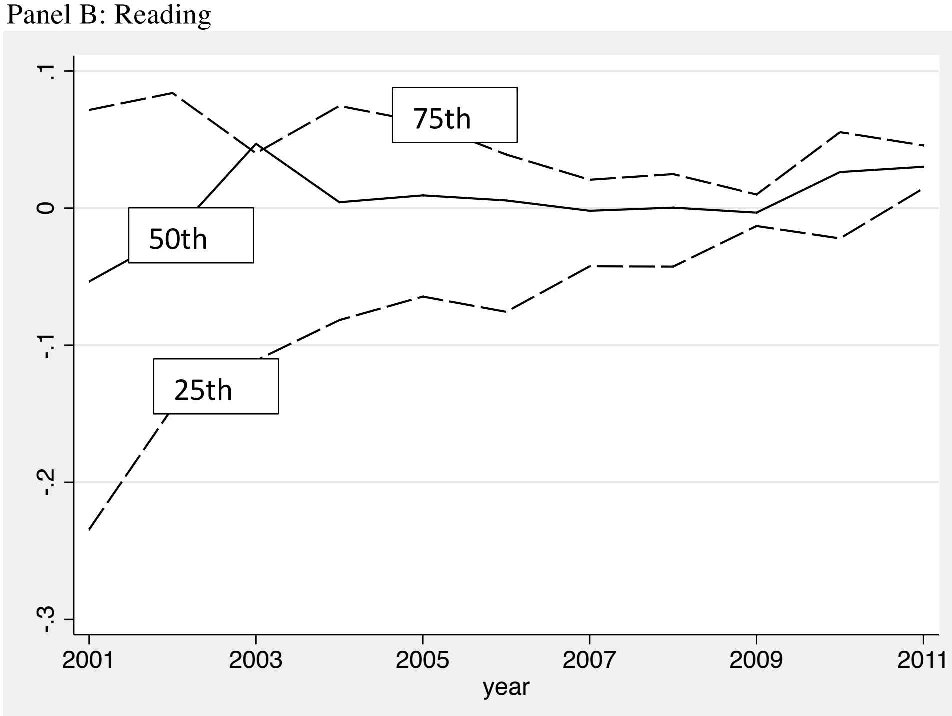
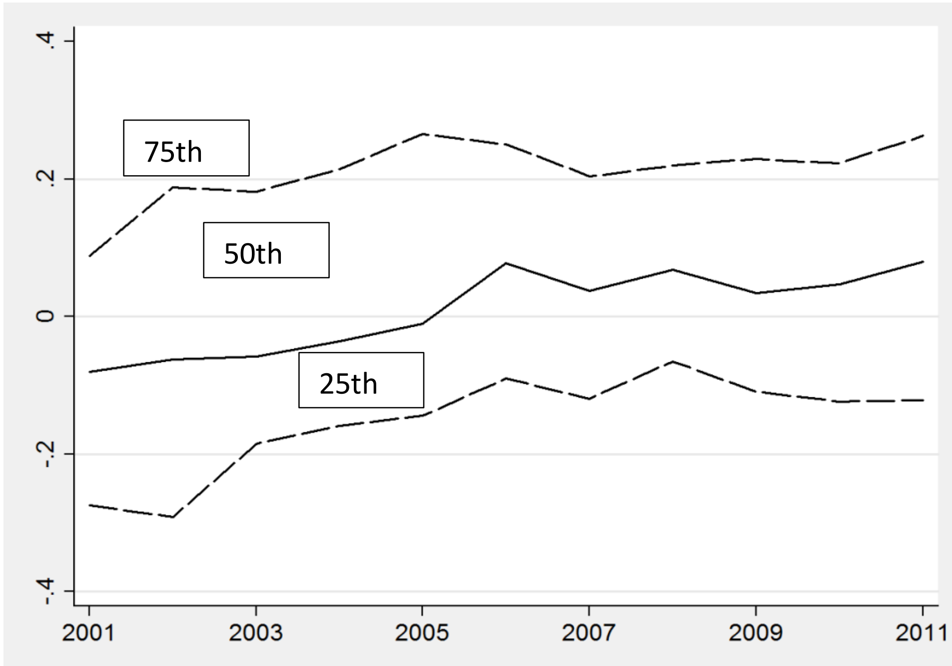


Figure 10: Charter School quality quartiles over time relative to TPS (Statewide Comparisons)

Note: Figures show the difference between the 25th, 50th, and 75th percentile of charter school quality distributions and the same percentile from the distributions of TPS quality based on statewide value-added models.

Panel A. Mathematics



Panel A. Reading

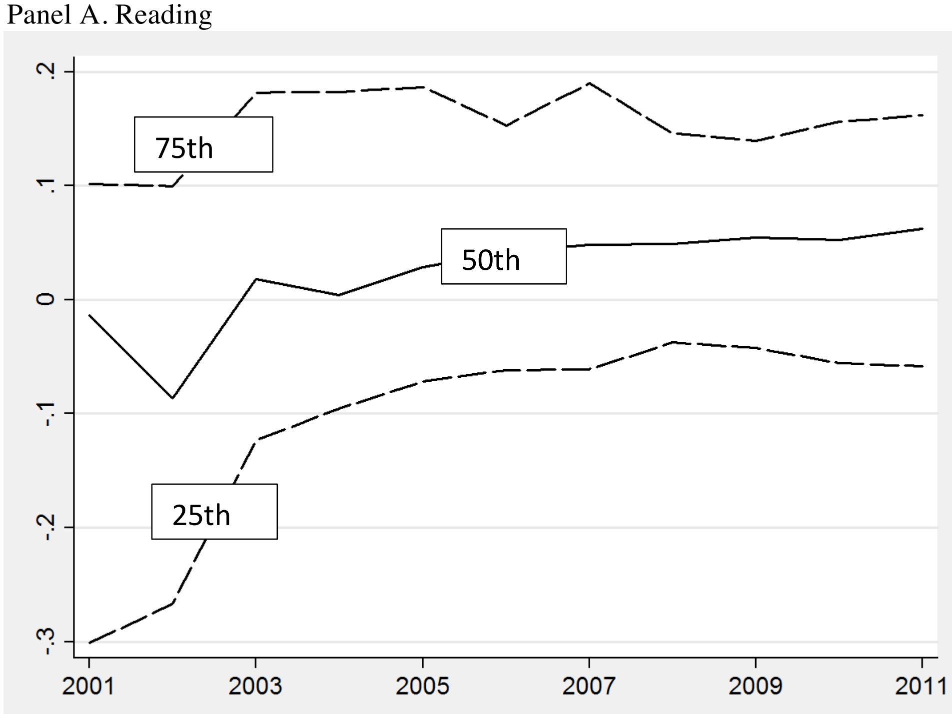


Figure 11. Charter School Quality Quartiles over time relative to TPS using Matching Procedure

Note: Figures show estimates at the 25th, 50th, and 75th percentile of the charter school quality distributions based on matching models.

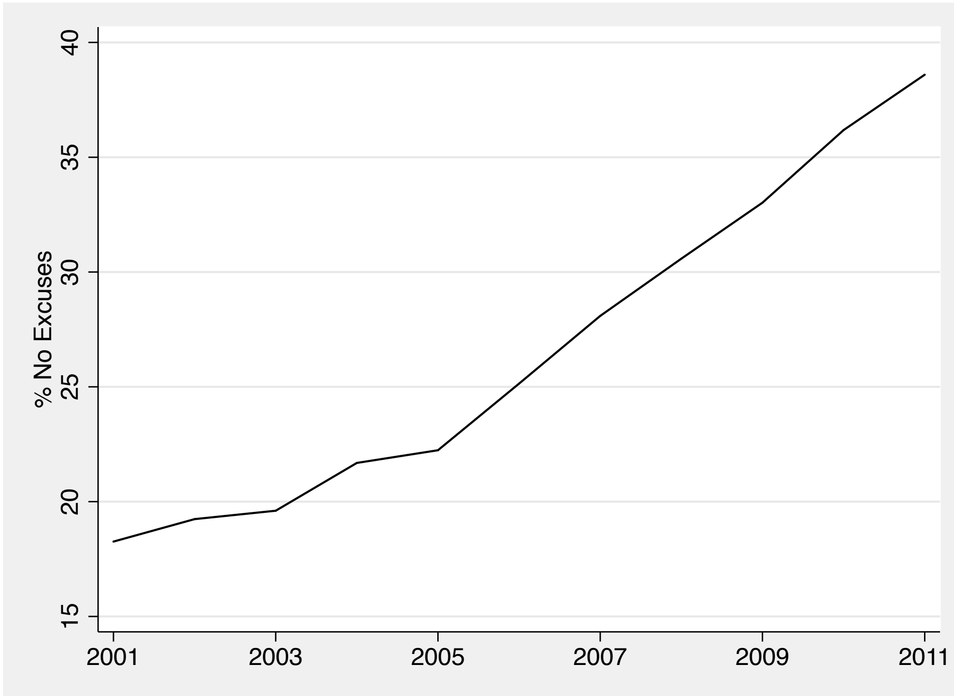


Figure 12: Trends over Time in the Share of Schools that Adhere to a No Excuses Philosophy

Note: No Excuses status is defined at the CMO level, and the percentage is expressed in terms of the number of students enrolled at a 'No Excuses' campus relative to all charter school students.

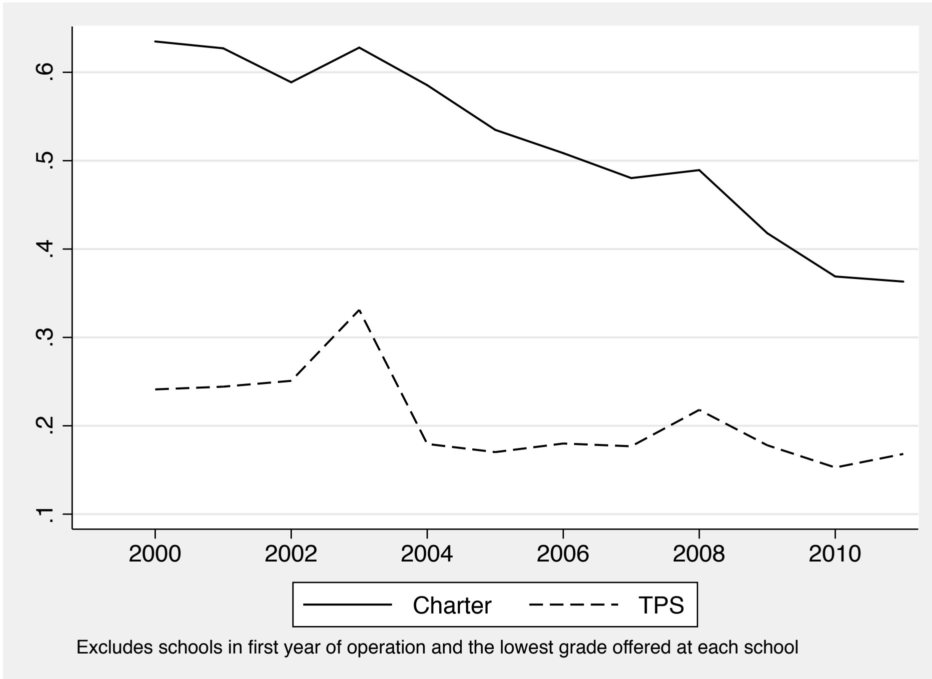
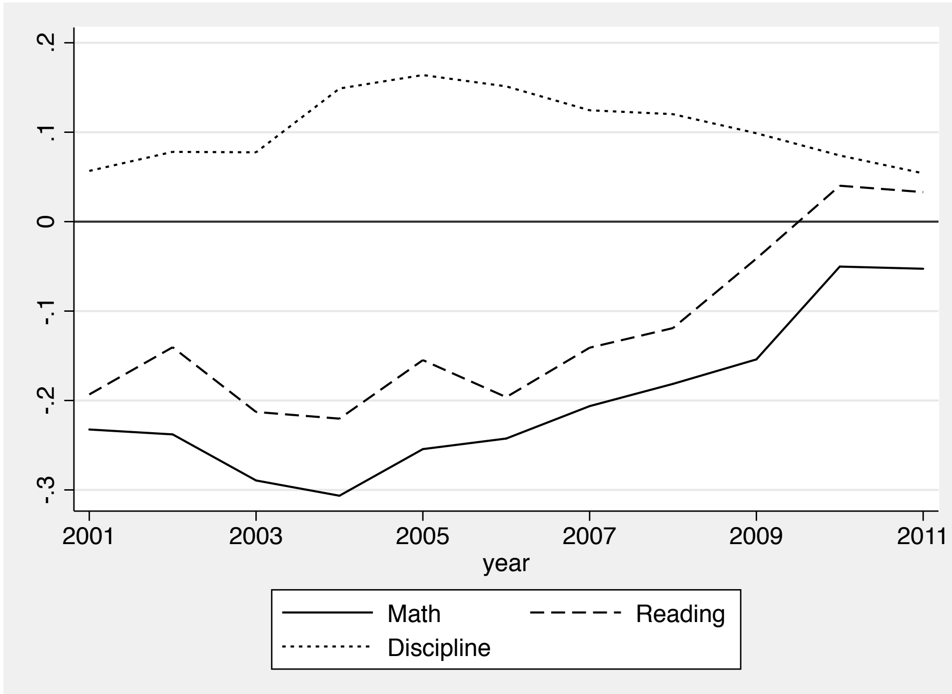


Figure 13: Proportion of Students that are New to the School in the Charter and Traditional Public School Sectors: 2001 to 2011

Panel A: All Charter School Entrants



Panel B: Charter Entrants Who Remain into Their Second Year

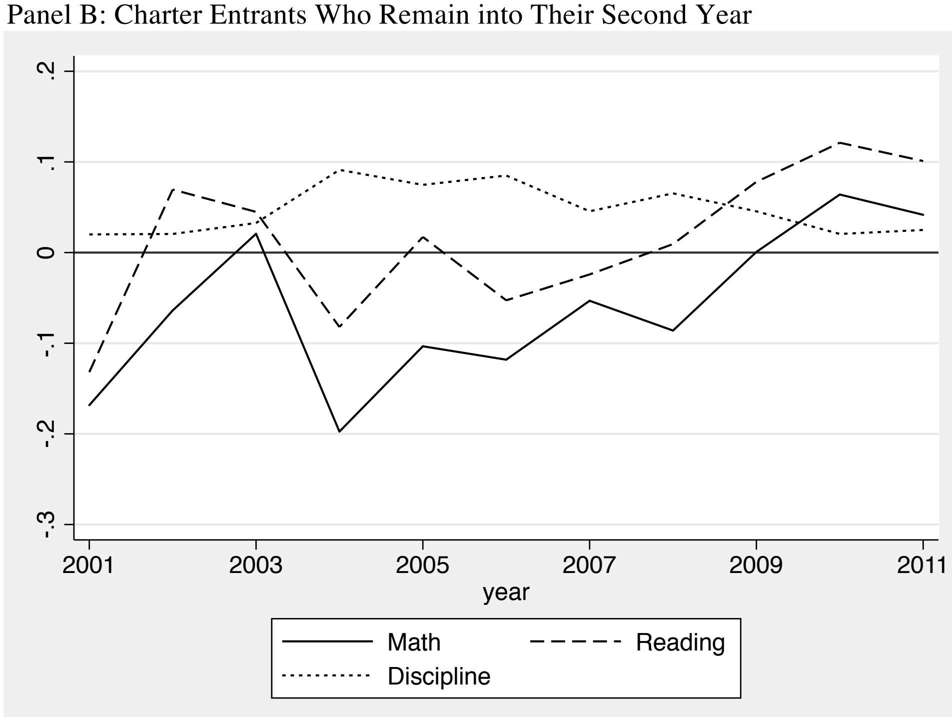


Figure 14: Trends Over Time in Selection into the Charter Sector by Prior Mathematics and Reading Achievement and the Probability of Receiving a Disciplinary Infraction: 2001-2011

Notes: These series compare students who transition to a charter school to their former schoolmates who remain at a TPS using information from the year prior to the transition. Math & reading refer to average achievement. Discipline refers to the probability of having committed any disciplinary infraction.

TABLES

Table 1. Descriptive Statistics by Rating

	Proportion of Schools	Enrollment Weighted	Proportion White	Black	Hispanic	Low Income	Average Enrollment
Exceptional	0.16	0.15	0.57	0.08	0.29	0.35	541
Recognized	0.31	0.33	0.41	0.11	0.44	0.53	586
Acceptable	0.38	0.45	0.32	0.16	0.48	0.59	702
Unacceptable	0.02	0.03	0.18	0.30	0.50	0.71	616
Unrated	0.08	0.02	0.22	0.21	0.54	0.75	109
AEA	0.05	0.01	0.23	0.22	0.52	0.62	107
Total			0.37	0.14	0.44	0.53	558

Notes: Demographic proportions are enrollment weighted. Average enrollment is a simple average without weighting by enrollment.

Table 2. School pass rate changes by Rating and Prior Rating

Panel A. % Point change in Math Pass Rate

		Rating: Exceptional	Recognized	Acceptable	Unacceptable	Total
Prior Rating	Exceptional	0.36	-1.65	-5.08	-13.74	-0.52
	Recognized	2.73	0.77	-2.21	-8.22	0.24
	Acceptable	7.23	4.53	1.66	-1.97	2.42
	Unacceptable	28.91	13.73	9.22	4.76	8.73
	Total	1.63	1.76	0.84	-1.27	1.27

Panel B. % Point change in Reading Pass Rate

		Rating: Exceptional	Recognized	Acceptable	Unacceptable	Total
Prior Rating	Exceptional	0.03	-1.56	-3.18	-12.86	-0.63
	Recognized	1.57	0.25	-1.5	-5.47	-0.04
	Acceptable	4.71	2.43	1.19	-1.12	1.51
	Unacceptable	16.96	6.95	4.47	2.51	4.32
	Total	0.88	0.78	0.56	-0.83	0.66

Notes: Each cell gives the average pass rate gain in percentage points among schools that received the rating down the left in the first year and the rating along the top in the second year.

Table 3. Effects of Prior Ratings on the percent of students that pass the Math test at Traditional Public Schools

	Lagged Pass Rate - OLS		Lagged Pass Rate - IV		Pass Rate Gain	
	(1)	(2)	(3)	(4)	(5)	(6)
Lagged Rating						
Exceptional	-0.824	-2.56	-1.091	-2.338	-3.063	-4.763
	(0.078)	(0.099)	(0.094)	(0.122)	(0.058)	(0.091)
Recognized	-0.276	-0.734	-0.627	-0.889	-1.922	-2.593
	(0.059)	(0.067)	(0.071)	(0.083)	(0.048)	(0.065)
Unacceptable	2.195	2.54	2.218	2.241	4.688	5.379
	(0.230)	(0.238)	(0.240)	(0.256)	(0.226)	(0.260)
Campus FE	N	Y	N	Y	N	Y
Mean	82.64	82.64	83.58	83.58	1.352	1.352
Std Dev	12.86	12.86	12.06	12.06	5.539	5.539
First Stage F	.	.	3609.8	1100	.	.
N	59561	59561	51303	51010	60300	60300

Notes: Each cell gives the coefficient on the specified prior rating dummy from a regression of math pass rates on prior rating dummies with *acceptable* the excluded group. All regressions are enrollment weighted and include year dummies and campus demographics. Columns (1) and (2) control for the lagged campus pass rates in both math and reading and columns (3) and (4) instrument for lagged pass rates with twice lagged pass rates. Columns (5) and (6) have the subsequent pass rate gain as the dependent variable. Even numbered columns control for campus fixed effects. Clustered standard errors at the campus level in parentheses.

Table 4. Effects of Prior Ratings on the percent of students that pass the Reading test at Traditional Public Schools

	Lagged Pass Rate - OLS		Lagged Pass Rate - IV		Pass Rate Gain	
	(1)	(2)	(3)	(4)	(5)	(6)
Lagged Rating						
Exceptional	0.196	-0.359	0.417	0.0849	-1.454	-2.057
	(0.056)	(0.068)	(0.065)	(0.083)	(0.043)	(0.069)
Recognized	0.122	-0.0719	0.175	0.107	-1.044	-1.188
	(0.042)	(0.046)	(0.049)	(0.056)	(0.036)	(0.049)
Unacceptable	0.535	1.121	0.221	0.291	2.301	2.191
	(0.160)	(0.172)	(0.168)	(0.176)	(0.178)	(0.220)
Campus FE	N	Y	N	Y	N	Y
Mean	88.93	88.93	89.51	89.51	0.761	0.761
Std Dev	7.584	7.584	6.999	6.999	4.4	4.4
First Stage F	.	.	3578.2	1090.2	.	.
N	59251	59251	51071	50778	59451	59451

Notes: Each cell gives the coefficient on the specified prior rating dummy from a regression of reading pass rates on prior rating dummies with *acceptable* the excluded group. All regressions are enrollment weighted and include year dummies and campus demographics. Columns (1) and (2) control for the lagged campus pass rates in both math and reading and columns (3) and (4) instrument for lagged pass rates with twice lagged pass rates. Columns (5) and (6) have the subsequent pass rate gain as the dependent variable. Even numbered columns control for campus fixed effects. Clustered standard errors at the campus level in parentheses.

Table 5. Effects of Prior Ratings on the proportion of students who reenroll at Traditional Public Schools

	1 Year		2 Year	
	(1)	(2)	(3)	(4)
Exceptional	0.001	0.01	0.009	0.012
	(0.003)	(0.002)	(0.005)	(0.004)
Recognized	-0.004	0.004	-0.002	0.005
	(0.002)	(0.001)	(0.003)	(0.002)
Unacceptable	-0.04	-0.008	-0.059	-0.016
	(0.006)	(0.004)	(0.009)	(0.005)
Campus FE	N	Y	N	Y
Mean	0.787		0.65	
StdDev	0.119		0.154	
N	45966		30072	

Notes: Each cell gives the coefficient on the specified prior rating dummy from a regression of the eligible student reenrollment proportion on prior rating dummies with *acceptable* the excluded group. All regressions are enrollment weighted and include year dummies and control for campus demographics and pass rates. Columns labeled "1 year" have the one year ahead reenrollment proportion as the dependent variable, and Columns labeled "2 year" have the two year ahead reenrollment proportion as the dependent variable. Even numbered columns control for campus fixed effects. Clustered standard errors at the campus level in parentheses.

Table 6. Regression Discontinuity estimates of the effect of receiving the lower of two neighboring ratings on the school pass rate on standardized tests the following year at Traditional Public Schools

Panel A. Math

	Group Size Discontinuity			Pass Rate Discontonuity		
	E to R	R to A	A to U	E to R	R to A	A to U
Reduced Form	-0.254	0.522	2.913	-0.253	0.649	0.788
	(0.567)	(0.541)	(1.282)	(0.225)	(0.307)	(0.583)
IV	-0.319	0.769	7.729	-0.421	1.080	9.132
	(0.711)	(0.797)	(3.485)	(0.374)	(0.512)	(6.807)
N	1237	2326	790	12391	12482	6125

Panel B. Reading

	Group Size Discontinuity			Pass Rate Discontonuity		
	E to R	R to A	A to U	E to R	R to A	A to U
Reduced Form	-0.298	0.824	0.255	-0.566	0.191	0.677
	(0.416)	(0.537)	(0.941)	(0.203)	(0.254)	(0.523)
IV	-0.367	1.248	0.728	-0.866	0.318	10.401
	(0.512)	(0.820)	(2.695)	(0.310)	(0.425)	(9.084)
N	1493	1846	577	12004	12342	5182

Notes: Effect of a negative rating shock on the pass rate the the next year measured in percentage points. The first three columns instrument for the rating using the size of the largest underperforming group. The second three columns instrument for rating using the pass-rate of the lowest performing group that is large enough to matter. All specifications control for the lagged pass rate. *E to R* columns give the pass rate effect of being rated *recognized* rather than *exemplary*, *R to A* columns give the pass rate effect of being rated *acceptable* rather than *recognized*, and *A to U* columns give the pass rate effect of being rated *unacceptable* rather than *acceptable*. I control for the distance from the RD threshold non-parametrically using local linear regression (LLR) with triangular kernel weights. Optimal bandwidths are selected using the data driven approach of Imbens and Kalyanaraman (2012). Clustered standard errors at the campus level in parentheses.

Table 7. Regression Discontinuity estimates of the effect of receiving the lower of two neighboring ratings on the school reenrollment rate the following year at Traditional Public Schools

Panel A. One Year Later

	Group Size Discontinuity			Pass Rate Discontonuity		
	E to R	R to A	A to U	E to R	R to A	A to U
Reduced Form	0.001	0.005	-0.005	-0.008	0.002	-0.009
	(0.015)	(0.012)	(0.034)	(0.005)	(0.005)	(0.009)
IV	0.002	0.007	-0.025	-0.015	0.003	-0.084
	(0.019)	(0.017)	(0.159)	(0.009)	(0.008)	(0.082)
N	370	612	277	4205	5563	1394

Panel B. Two Years Later

	Group Size Discontinuity			Pass Rate Discontonuity		
	E to R	R to A	A to U	E to R	R to A	A to U
Reduced Form	-0.010	0.016	-0.005	-0.020	0.007	0.000
	(0.042)	(0.018)	(0.053)	(0.008)	(0.007)	(0.012)
IV	-0.014	0.024	-0.024	-0.035	0.014	-0.003
	(0.058)	(0.028)	(0.274)	(0.015)	(0.014)	(0.109)
N	278	459	208	3154	4172	1046

Notes: Effect of a negative rating shock on the proportion of students who reenroll the next and subsequent years. The first three columns instrument for the rating using the size of the largest underperforming group. The second three columns instrument for rating using the pass-rate of the lowest performing group that is large enough to matter. *E to R* columns give the reenrollment rate effect of being rated *recognized* rather than *exemplary*, *R to A* columns give the reenrollment rate effect of being rated *acceptable* rather than *recognized*, and *A to U* columns give the reenrollment rate effect of being rated *unacceptable* rather than *acceptable*. I control for the distance from the RD threshold non-parametrically using local linear regression (LLR) with triangular kernel weights. Optimal bandwidths are selected using the data driven approach of Imbens and Kalyanaraman (2012). Clustered standard errors at the campus level in parentheses.

Table 8. Effects of Prior Ratings on the percent of students that pass the Math test at Charter Schools

	Lagged Pass Rate - OLS		Lagged Pass Rate - IV		Pass Rate Gain	
	(1)	(2)	(3)	(4)	(5)	(6)
Lagged Rating						
Exceptional	1.363	-3.441	-2.962	7.305	-6.301	-13.000
	(1.065)	(1.457)	(2.098)	(18.650)	(0.895)	(1.935)
Recognized	1.111	-1.263	-1.425	6.636	-4.160	-7.880
	(0.990)	(1.061)	(1.440)	(13.150)	(0.953)	(1.546)
Unacceptable	0.870	2.327	5.561	-5.008	7.492	11.680
	(1.291)	(1.395)	(1.933)	(12.060)	(1.249)	(1.746)
Campus FE	N	Y	N	Y	N	Y
Mean	77.16	77.16	77.99	77.99	2.734	2.734
Std Dev	17.53	17.53	16.29	16.29	11.79	11.79
First Stage F	.	.	25.47	0.201	.	.
N	1222	1222	906	825	1263	1263

***Notes*: Each cell gives the coefficient on the specified prior rating dummy from a regression of math pass rates on prior rating dummies with *acceptable* the excluded group. All regressions are enrollment weighted and include year dummies and campus demographics. Columns (1) and (2) control for the lagged campus pass rates in both math and reading and columns (3) and (4) instrument for lagged pass rates with twice lagged pass rates. Columns (5) and (6) have the subsequent pass rate gain as the dependent variable. Even numbered columns control for campus fixed effects. Clustered standard errors at the campus level in parentheses.**

Table 9. Effects of Prior Ratings on the percent of students that pass the Reading test at Charter Schools

	Lagged Pass Rate - OLS		Lagged Pass Rate - IV		Pass Rate Gain	
	(1)	(2)	(3)	(4)	(5)	(6)
Lagged Rating						
Exceptional	1.403	-1.849	0.55	-0.49	-1.668	-5.058
	(0.785)	(1.068)	(1.525)	(5.527)	(0.570)	(1.194)
Recognized	0.244	-1.504	-0.213	-0.424	-1.693	-3.639
	(0.565)	(0.645)	(1.040)	(4.313)	(0.507)	(0.817)
Unacceptable	-0.332	1.123	2.486	-0.22	4.464	6.056
	(0.962)	(0.970)	(1.716)	(3.436)	(1.026)	(1.489)
Campus FE	N	Y	N	Y	N	Y
Mean	86.01	86.01	86.46	86.46	1.17	1.17
Std Dev	11.10	11.10	10.27	10.27	8.30	8.30
First Stage F	.	.	24.57	0.516	.	
N	1212	1212	902	824	1233	1233

Notes: Each cell gives the coefficient on the specified prior rating dummy from a regression of reading pass rates on prior rating dummies with *acceptable* the excluded group. All regressions are enrollment weighted and include year dummies and campus demographics. Columns (1) and (2) control for the lagged campus pass rates in both math and reading and columns (3) and (4) instrument for lagged pass rates with twice lagged pass rates. Columns (5) and (6) have the subsequent pass rate gain as the dependent variable. Even numbered columns control for campus fixed effects. Clustered standard errors at the campus level in parentheses.

Table 10. Effects of Prior Ratings on the proportion of students who reenroll at Charter Schools

	1 Year		2 Year	
	(1)	(2)	(3)	(4)
Exceptional	0.009	-0.003	0.042	-0.01
	(0.026)	(0.025)	(0.040)	(0.040)
Recognized	0.017	0.012	0.035	0.004
	(0.016)	(0.016)	(0.020)	(0.014)
Unacceptable	-0.065	-0.031	-0.078	-0.029
	(0.020)	(0.018)	(0.026)	(0.017)
Campus FE	N	Y	N	Y
Mean	0.68		0.512	
StdDev	0.196		0.224	
N	1173		853	

Notes: Each cell gives the coefficient on the specified prior rating dummy from a regression of the eligible student reenrollment proportion on prior rating dummies with *acceptable* the excluded group. All regressions are enrollment weighted and include year dummies and control for campus demographics and pass rates. Columns labeled "1 year" have the one year ahead reenrollment proportion as the dependent variable, and Columns labeled "2 year" have the two year ahead reenrollment proportion as the dependent variable. Even numbered columns control for campus fixed effects. Clustered standard errors at the campus level in parentheses.

Table 11. Regression Discontinuity estimates of the effect of receiving the lower of two neighboring ratings on the school pass rate on standardized tests the following year at Charter Schools

Panel A. Math

	Group Size Discontinuity			Pass Rate Discontonuity		
	E to R	R to A	A to U	E to R	R to A	A to U
Reduced Form	-1.858	-15.190	3.720	0.298	2.270	6.674
	(3.323)	(2.614)	(5.356)	(2.359)	(3.462)	(4.111)
IV	-2.756	-21.684	10.556	0.846	4.267	19.309
	(4.822)	(4.481)	(15.569)	(6.700)	(6.603)	(12.887)
N	109	126	140	339	445	670

Panel B. Reading

	Group Size Discontinuity			Pass Rate Discontonuity		
	E to R	R to A	A to U	E to R	R to A	A to U
Reduced Form	-0.641	-6.062	0.454	-1.997	3.867	0.508
	(2.258)	(2.407)	(2.469)	(2.111)	(2.687)	(3.255)
IV	-0.779	-9.220	1.899	-3.323	6.780	1.865
	(2.752)	(2.635)	(10.006)	(3.547)	(5.017)	(11.765)
N	150	88	107	348	366	670

Notes: Effect of a negative rating shock on the pass rate the the next year measured in percentage points. The first three columns instrument for the rating using the size of the largest underperforming group. The second three columns instrument for rating using the pass-rate of the lowest performing group that is large enough to matter. All specifications control for the lagged pass rate. *E to R* columns give the pass rate effect of being rated *recognized* rather than *exemplary*, *R to A* columns give the pass rate effect of being rated *acceptable* rather than *recognized*, and *A to U* columns give the pass rate effect of being rated *unacceptable* rather than *acceptable*. I control for the distance from the RD threshold non-parametrically using local linear regression (LLR) with triangular kernel weights. Optimal bandwidths are selected using the data driven approach of Imbens and Kalyanaraman (2012). Clustered standard errors at the campus level in parentheses.

Table 12. Regression Discontinuity estimates of the effect of receiving the lower of two neighboring ratings on the school reenrollment rate the following year at Charter Schools

Panel A. One Year Later

	Group Size Discontinuity			Pass Rate Discontonuity		
	E to R	R to A	A to U	E to R	R to A	A to U
Reduced Form	-0.163	0.130	0.014	0.120	-0.458	-0.094
	(0.123)	(0.064)	(0.019)	(0.063)	(0.207)	(0.171)
IV	-0.490	0.130	0.060	0.279	-0.518	-0.157
	(0.767)	(0.065)	(0.096)	(0.192)	(0.248)	(0.306)
N	84	120	171	261	424	818

Panel B. Two Years Later

	Group Size Discontinuity			Pass Rate Discontonuity		
	E to R	R to A	A to U	E to R	R to A	A to U
Reduced Form	-	0.270	0.110	0.194	-0.307	-0.201
	-	0.201	0.043	0.118	0.171	0.140
IV	-	0.358	0.985	4.446	-0.508	-1.163
	-	(0.275)	(2.075)	(27.427)	(0.286)	(3.370)
N	-	115	143	245	389	738

Notes: Effect of a negative rating shock on the proportion of students who reenroll the next and subsequent years. The first three columns instrument for the rating using the size of the largest underperforming group. The second three columns instrument for rating using the pass-rate of the lowest performing group that is large enough to matter. *E to R* columns give the reenrollment rate effect of being rated *recognized* rather than *exemplary*, *R to A* columns give the reenrollment rate effect of being rated *acceptable* rather than *recognized*, and *A to U* columns give the reenrollment rate effect of being rated *unacceptable* rather than *acceptable*. I control for the distance from the RD threshold non-parametrically using local linear regression (LLR) with triangular kernel weights. Optimal bandwidths are selected using the data driven approach of Imbens and Kalyanaraman (2012). Clustered standard errors at the campus level in parentheses.

Table 13. Average Charter School Mathematics and Reading Value-added and Enrollment Shares for 2001 and 2011, by status of school operations (Statewide estimates)

	Mathematics		Reading	
	2001	2011	2001	2011
A. Schools in operation in 2001 and in 2011				
Average Value-added	-0.09	0.01	-0.03	-0.01
Share of Charter Enrollment	0.69	0.26	0.69	0.26
Number of Schools	96		96	
B. Market Closures				
Average Value-added	-0.25	.	-0.22	.
Share of Charter Enrollment	0.20	.	0.20	.
Number of Schools	45		45	
C. Authorizer Closures				
Average Value-added	-0.16	.	-0.23	.
Share of Charter Enrollment	0.11	.	0.11	.
Number of Schools	9		9	
D. Schools in operation in 2011 but not in 2001				
Average Value-added	.	-0.02	.	0.01
Share of Charter Enrollment	.	0.74	.	0.74
Number of Schools	320		319	

Notes: Average value-added for charter schools weighted by enrollment; traditional public school average value-added in each year deducted from the corresponding charter average. Empty cells in panels B, C and D correspond to years when these school categories are no longer in operation or have yet to begin operation. Estimates are constructed using statewide comparison group.

Table 14: Estimated Effects of Prior Year CMO Performance on the Number of Schools Operated (Statewide estimates)

	Net Change		Net Expansion		Net Contraction	
	(1)	(2)	(3)	(4)	(5)	(6)
CMO Average Math VA	0.168	0.150	0.047	0.038	-0.061	-0.057
	(0.068)	(0.074)	(0.026)	(0.028)	(0.018)	(0.020)
CMO Average Reading VA	0.229	0.203	0.036	0.021	-0.107	-0.092
	(0.084)	(0.090)	(0.032)	(0.033)	(0.022)	(0.023)
CMO FE	No	Yes	No	Yes	No	Yes
Mean	0.139		0.120		0.055	
N	1847		1847		1847	

Note: Data for regressions include all CMOs operating in each year. Each estimate comes from a separate regression. All regressions include year dummies. The dependent variable in columns (1) and (2) is the net change in the number of campuses in operation for a CMO, while in columns (3) - (6) the dependent variable is an indicator equal to one if the net change is positive (expansion) or negative (contraction). Estimates are constructed using statewide comparison group. Standard errors are clustered at the CMO level.

Table 15. Average Charter School Mathematics and Reading Value-added and Enrollment Shares for 2001 and 2011, by Status of School Operations (Matching estimates)

	Mathematics		Reading	
	2001	2011	2001	2011
A. Schools in operation in 2001 and in 2011				
Average Value-added	-0.01	0.05	-0.03	0.00
Share of Charter Enrollment	0.68	0.26	0.68	0.26
Number of Schools	92		92	
B. Market Closures				
Average Value-added	-0.17	.	-0.17	.
Share of Charter Enrollment	0.22	.	0.22	.
Number of Schools	44		44	
C. Authorizer Closures				
Average Value-added	-0.05	.	-0.34	.
Share of Charter Enrollment	0.10	.	0.10	.
Number of Schools	9		9	
D. Schools in operation in 2011 but not in 2001				
Average Value-added	.	0.09	.	0.07
Share of Charter Enrollment	.	0.74	.	0.74
Number of Schools	319		319	

Notes: Average value-added for charter schools is weighted by enrollment;. Empty cells in panels B, C and D correspond to years when these school categories are no longer in operation or have yet to begin operation. Estimates are constructed using matching model comparison group.

Table 16. Estimated Effects of Prior Year CMO Performance on the Number of Schools Operated (Matching estimates)

	Net Change		Net Expansion		Net Contraction	
	(1)	(2)	(3)	(4)	(5)	(6)
CMO Average Math VA	0.118	0.127	0.039	0.040	-0.029	-0.040
	(0.060)	(0.070)	(0.023)	(0.026)	(0.016)	(0.019)
CMO Average Reading VA	0.197	0.175	0.059	0.050	-0.061	-0.054
	(0.071)	(0.075)	(0.027)	(0.028)	(0.019)	(0.020)
CMO FE	No	Yes	No	Yes	No	Yes
Mean	0.139		0.120		0.055	
N	1847		1847		1847	

*Note*s: Data for regressions include all CMOs operating in each year. Each estimate comes from a separate regression. All regressions include year dummies. The dependent variable in columns (1) and (2) is the net change in the number of campuses in operation for a CMO, while columns in (3) - (6) the dependent variable is an indicator equal to one if the net change is positive (expansion) or negative (contraction). Estimates are constructed using matching model comparison group. Standard errors are clustered at the CMO level.

Table 17. Estimated Effects of Program Characteristics and Student Selection on Charter School Value-added (Statewide estimates)

	Mathematics				Reading			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
No Excuses Indicator	0.164	0.097	0.092	0.092	0.091	0.038	0.027	0.028
	(0.023)	(0.020)	(0.021)	(0.021)	(0.016)	(0.014)	(0.014)	(0.014)
Proportion New		-0.373	-0.337	-0.359		-0.297	-0.227	-0.242
		(0.035)	(0.041)	(0.045)		(0.031)	(0.036)	(0.041)
Achievement Difference								
Entrants			0.025	0.034			0.063	0.068
			(0.020)	(0.021)			(0.019)	(0.021)
Persisters			0.008	0.012			0.014	0.014
			(0.014)	(0.015)			(0.012)	(0.012)
Infraction Rate Difference								
Entrants				0.073				0.059
				(0.047)				(0.048)
Persisters				-0.006				-0.024
				(0.035)				(0.035)
N	1409				1396			

Note: The estimates come from school-by-year-level regressions with estimated value-added produced by the statewide comparison model as the dependent variable. Regressions include demographic characteristics and year dummies. Standard errors are clustered at the campus level.

Table 18. Estimated Effects of Program Characteristics and Student Selection on Charter School Value-added (Matching model estimates)

	Mathematics				Reading			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
No Excuses Indicator	0.185	0.128	0.127	0.129	0.113	0.067	0.061	0.066
	(0.030)	(0.029)	(0.030)	(0.030)	(0.022)	(0.023)	(0.023)	(0.022)
Proportion New		-0.317	-0.317	-0.360		-0.225	-0.194	-0.389
		(0.042)	(0.044)	(0.047)		(0.043)	(0.047)	(0.051)
Achievement Difference								
Entrants			-0.014	0.002			0.012	0.029
			(0.021)	(0.023)			(0.020)	(0.022)
Persisters			0.017	0.026			0.025	0.031
			(0.015)	(0.015)			(0.014)	(0.014)
Infraction Rate Difference								
Entrants				0.103				0.121
				(0.057)				(0.055)
Persisters				0.029				0.031
				(0.033)				(0.035)
N	1397				1385			

Note: The estimates come from school-by-year-level regressions with estimated value-added produced by the matching model as the dependent variable. Regressions include demographic characteristics and year dummies. Standard errors are clustered at the campus level.

Table 19. Estimated effects of Specific School Policies and Student Selection on Charter School Value-added (Statewide Estimates)

	Math		Reading	
	(1)	(2)	(3)	(4)
Uniforms	0.155	0.103	0.052	0.013
	(0.033)	(0.033)	(0.027)	(0.024)
Dress Code	0.046	0.034	0.010	0.007
	(0.031)	(0.029)	(0.026)	(0.023)
High Dosage Tutoring	-0.006	-0.015	-0.010	-0.024
	(0.032)	(0.031)	(0.021)	(0.022)
Parental Engagement	0.014	0.007	0.010	0.006
	(0.034)	(0.033)	(0.026)	(0.026)
High Expectations	0.050	0.033	0.034	0.020
	(0.027)	(0.026)	(0.020)	(0.019)
Days per year	0.005	0.006	-0.001	0.000
	(0.004)	(0.004)	(0.003)	(0.003)
Hours per year	-0.012	-0.027	0.008	-0.005
	(0.012)	(0.011)	(0.008)	(0.008)
Proportion New		-0.271		-0.194
		(0.046)		(0.041)
Achievement Difference				
Entrants		0.042		0.070
		(0.022)		(0.021)
Persisters		0.019		0.013
		(0.014)		(0.012)
Infraction Rate Difference				
Entrants		0.071		0.067
		(0.047)		(0.048)
Persisters		-0.019		-0.030
		(0.035)		(0.035)
N	1409	1409	1396	1396

Note: The estimates come from school-by-year-level regressions with estimated value-added produced by the statewide comparison model as the dependent variable. Regressions include demographic characteristics and year dummies. Standard errors are clustered at the campus level.

Table 20. Estimated effects of Specific School Policies and Student Selection on Charter School Value-added (Matching Model Estimates)

	Math		Reading	
	(1)	(2)	(3)	(4)
Uniforms	0.162	0.129	0.137	0.117
	(0.041)	(0.042)	(0.035)	(0.035)
Dress Code	0.063	0.054	0.081	0.077
	(0.037)	(0.036)	(0.031)	(0.030)
High Dosage Tutoring	-0.004	0.002	-0.001	0.000
	(0.047)	(0.045)	(0.034)	(0.032)
Parental Engagement	-0.026	-0.034	-0.023	-0.028
	(0.048)	(0.047)	(0.035)	(0.033)
High Expectations	0.062	0.053	0.045	0.042
	(0.033)	(0.032)	(0.027)	(0.026)
Days per year	0.002	0.002	0.000	0.000
	(0.005)	(0.005)	(0.003)	(0.003)
Hours per year	-0.011	-0.020	-0.013	-0.018
	(0.014)	(0.013)	(0.011)	(0.010)
Proportion New		-0.274		-0.167
		(0.055)		(0.048)
Achievement Difference				
Entrants		0.006		0.035
		(0.023)		(0.021)
Persisters		0.032		0.031
		(0.015)		(0.014)
Infraction Rate Difference				
Entrants		0.117		0.123
		(0.054)		(0.056)
Persisters		0.026		0.030
		(0.034)		(0.035)
N	1397	1397	1385	1385

Note: The estimates come from school-by-year-level regressions with estimated value-added produced by the matching model as the dependent variable. Regressions include demographic characteristics and year dummies. Standard errors are clustered at the campus level.

APPENDIX A - Classification of Schools as Adhering to a No Excuses Philosophy

We used a number of sources of information to determine whether a CMO adhered to the No Excuses philosophy. First, our research assistant called each school, described our project, and asked the representative if they could answer some questions about the school's approach to education. This often proved difficult, as many offered vague or curt responses. The research assistant then explored the website (if available), focusing on the mission or vision statements, superintendent's message, history, and other relevant information to gain a general feel for the school. Perhaps the most important source of information was the school handbook and code of conduct, and the research assistant carefully sifted through these documents. Finally, if none of these sources proved adequate, the research assistant searched for school reviews and articles that provided information on school policies and practices.

We focused on six areas to determine whether to classify a school as adhering to the No Excuses Philosophy. These areas are the following:

- **Discipline:** Most schools follow a progressive disciplinary system and provide clear expectations for behavior. Some schools, however, stand out as being particularly strict. We classify schools as strict in the discipline dimension if they use corporal punishment, impose strict zero tolerance policies for misbehavior, curfews, fine dining requirements (no talking or sharing), or sizable monetary fines for having cell phones or electronics, or undertake legal prosecution if a teacher is offended by students' language or other actions.
- **Expectations:** We use the following questions to determine whether a school sets very high expectations: Does the school hold all students to the same high expectations regardless of extraneous circumstances or family background? Does the school follow state standards or hold their students to higher expectations (i.e. are students required to meet state required 90

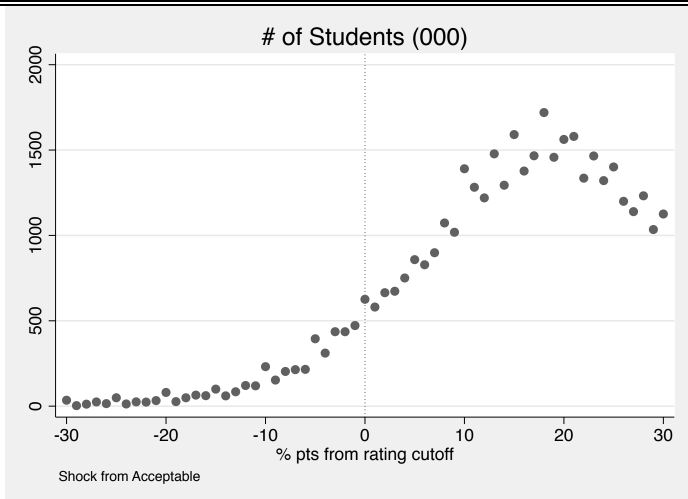
percent compulsory attendance or do they require *all* students to maintain 95-100 percent attendance to stay enrolled?)? Does the school require that all students are accepted at a university? Are students expected to graduate from college?

- **Uniforms**: Does the school require students to wear uniforms? Adhere to a strict dress code? Are there serious consequences for failing to comply? Are students sent home? Fined? Given detention? How many infractions until there is a serious consequence?
- **Parental Involvement:** Are parents encouraged to actively participate in the school? Are parents required to sign a commitment form?
- **Incentives**: Does the school offer rewards to students who surpass expectations? Most schools recognize students through things such as honor roll, by allowing them to go on field trips, or by letting them have a free dress day. Some offer additional incentives such as monetary prizes or privileges for good grades, attendance, and have a strong belief in reinforcing good behavior.
- **Extra**: Is there an extended school day? Week? Year? Is Saturday school offered or required? Tutoring?

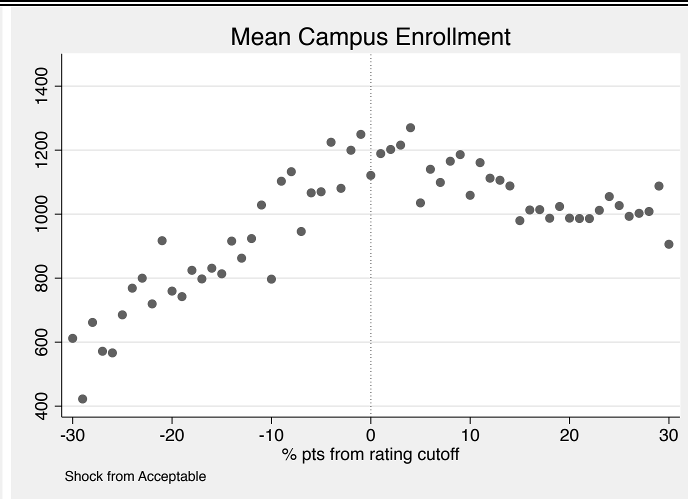
For some CMOs that were consistent across categories the classification decision was straightforward. For other CMOs the decision was more difficult, because they appeared to be strict in some dimensions but not others. In classifying these schools, we placed particular emphasis on the strictness of the disciplinary practices.

APPENDIX B – Tests of the Assumptions of RD

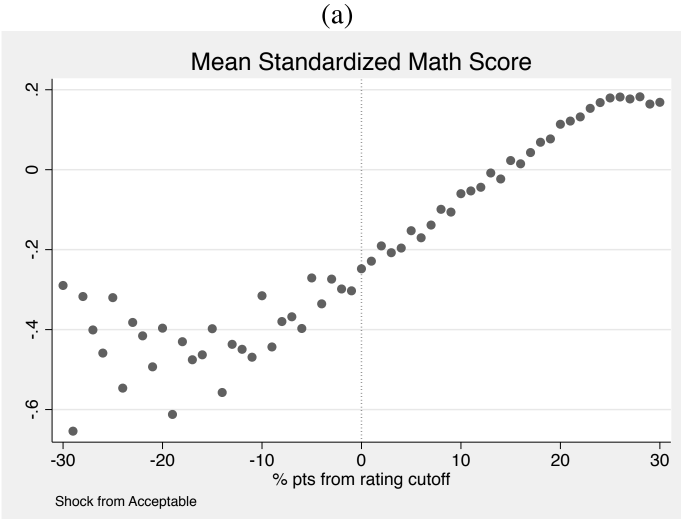
Figure 15. RD Assumptions: *Acceptable-Unacceptable* Pass-rate Boundary.



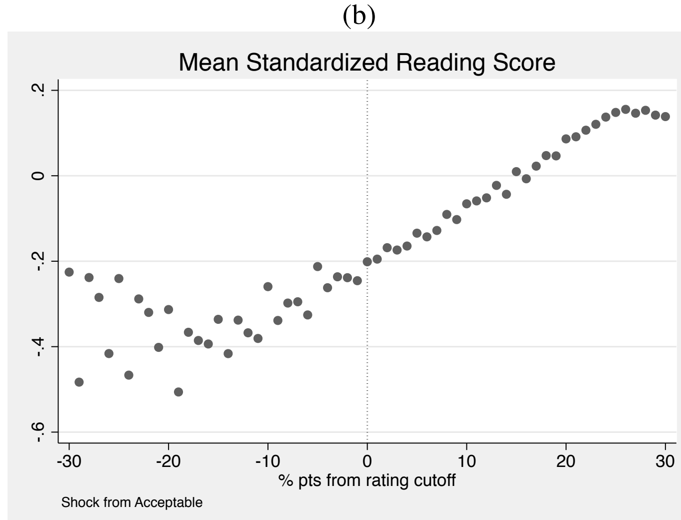
(a)



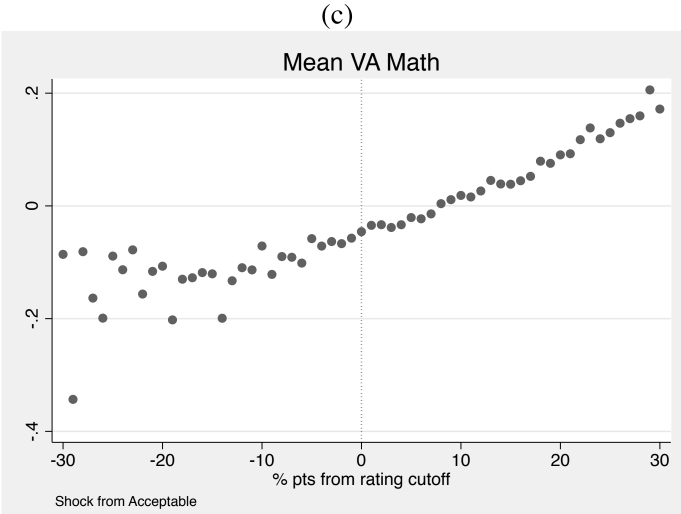
(b)



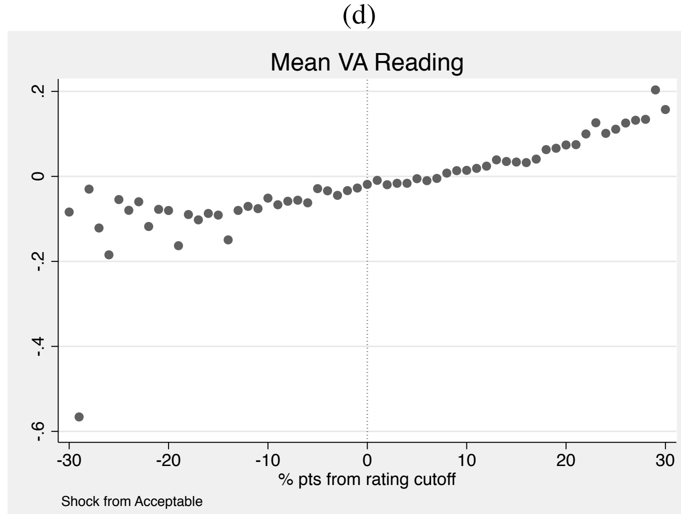
(c)



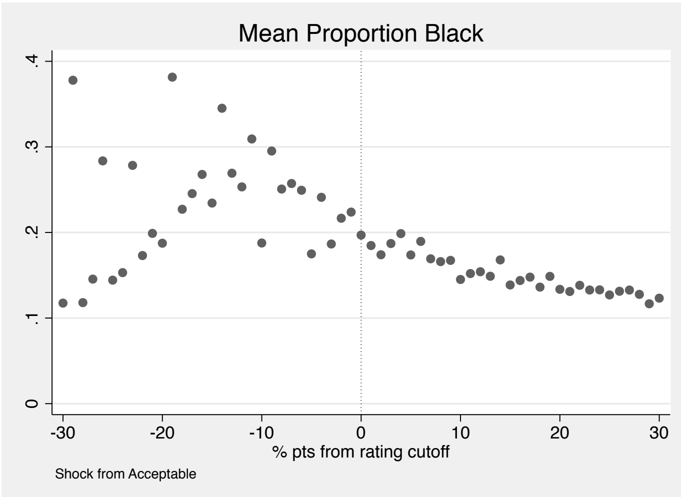
(d)



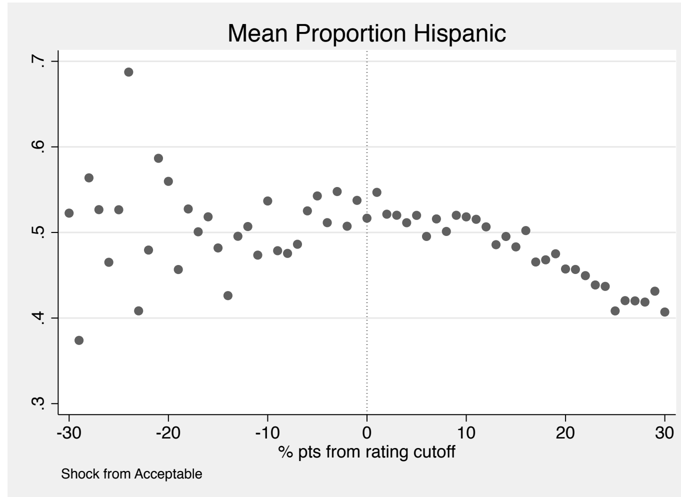
(e)



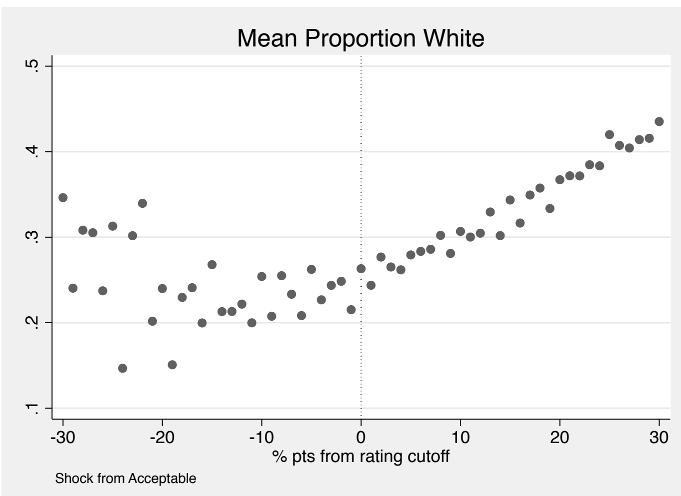
(f)


Mean Proportion Black
% pts from rating cutoff
Shock from Acceptable

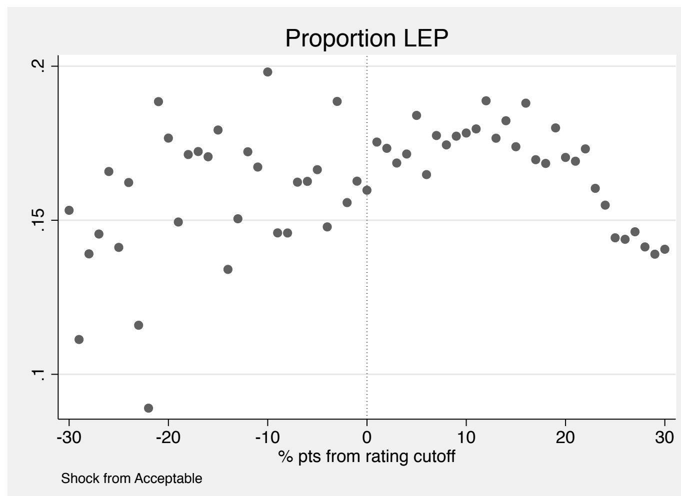
(g)


Mean Proportion Hispanic
% pts from rating cutoff
Shock from Acceptable

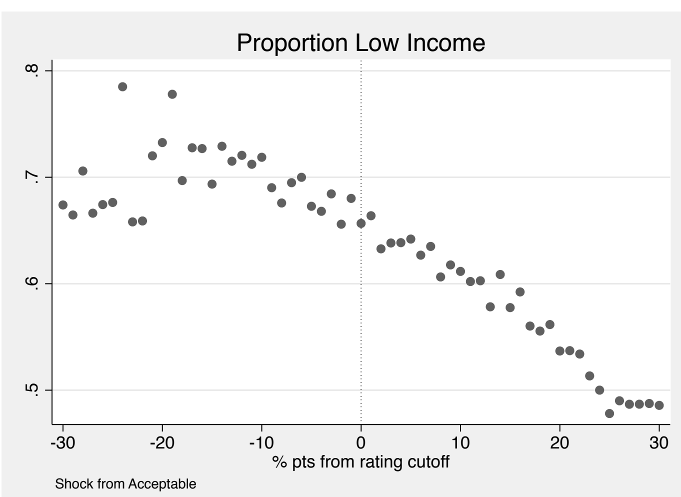
(h)


Mean Proportion White
% pts from rating cutoff
Shock from Acceptable

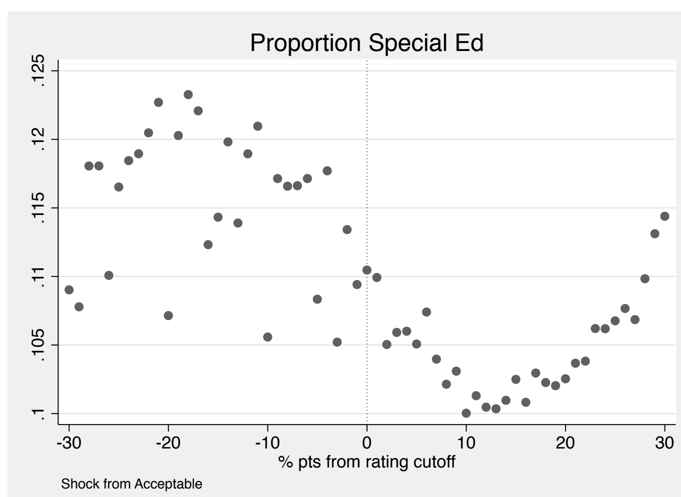
(i)


Proportion LEP
% pts from rating cutoff
Shock from Acceptable

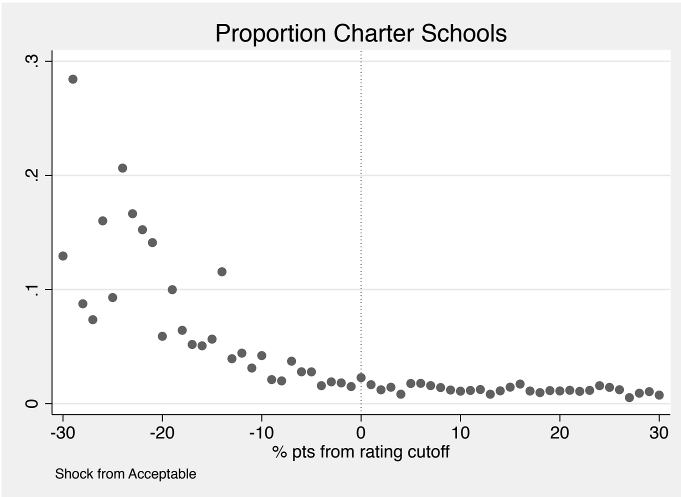
(j)


Proportion Low Income
% pts from rating cutoff
Shock from Acceptable

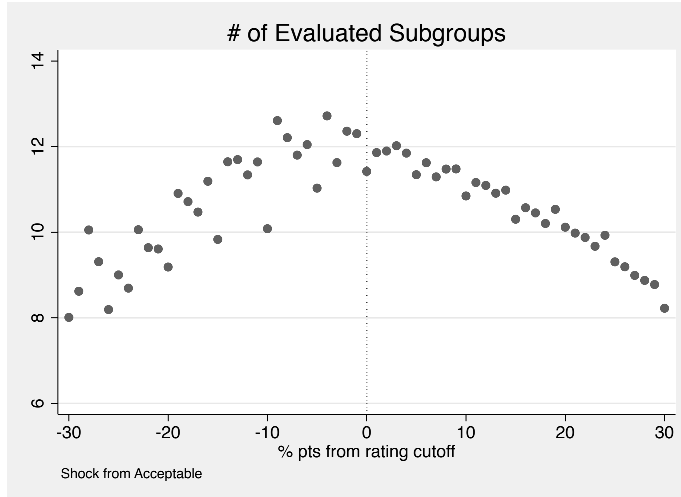
(k)


Proportion Special Ed
% pts from rating cutoff
Shock from Acceptable

(l)

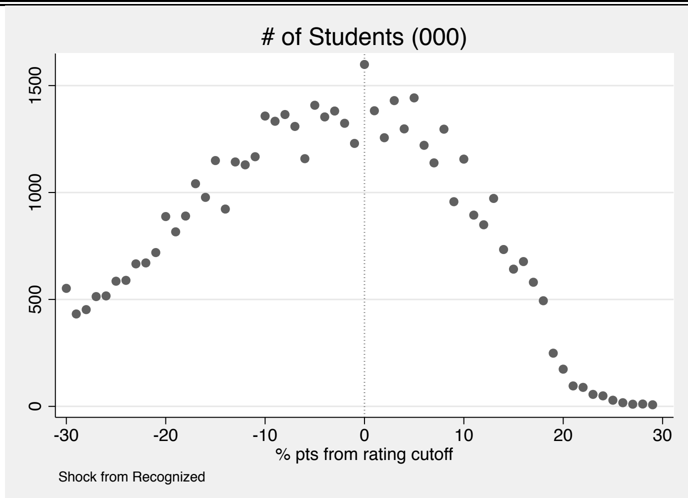

Proportion Charter Schools
0
.1
.2
.3
-30
-20
-10
0
10
20
30
% pts from rating cutoff
Shock from Acceptable

(m)

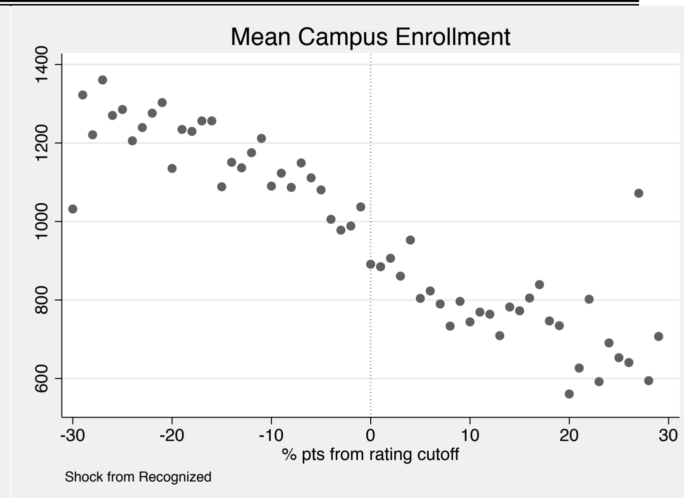

# of Evaluated Subgroups
6
8
10
12
14
-30
-20
-10
0
10
20
30
% pts from rating cutoff
Shock from Acceptable

(n)

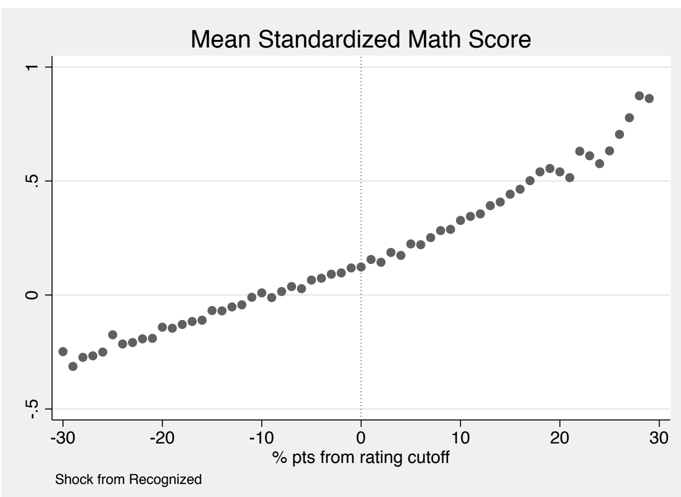
Figure 16. RD Assumptions: *Recognized-Acceptable* Pass-rate Boundary.



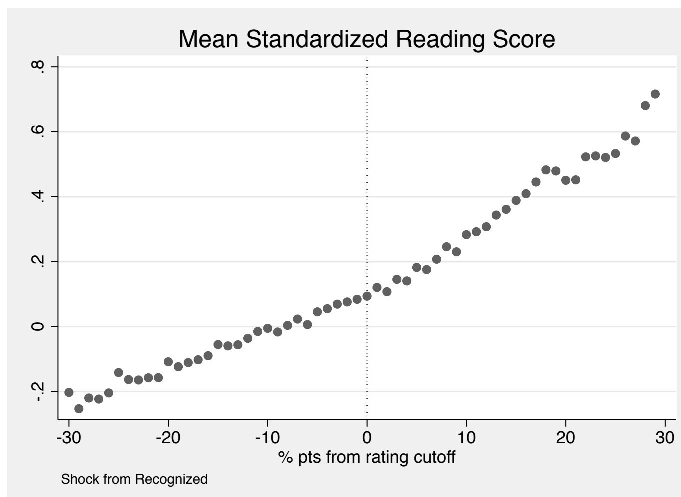
(a)



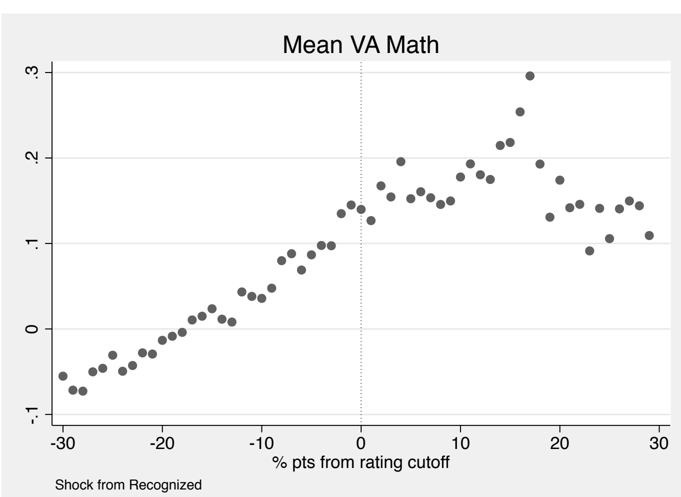
(b)



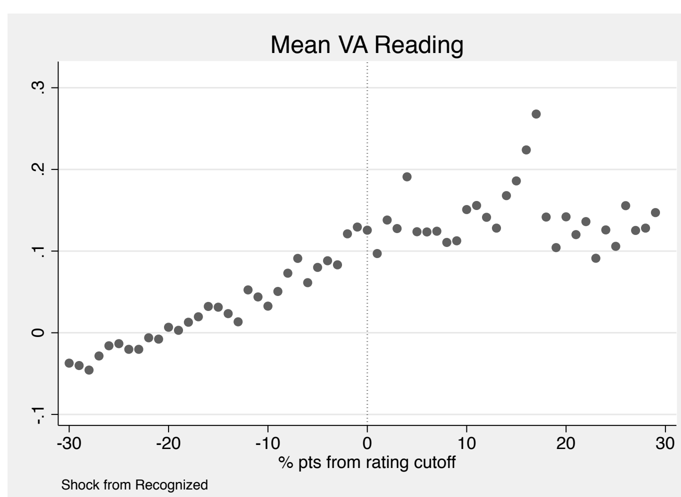
(c)



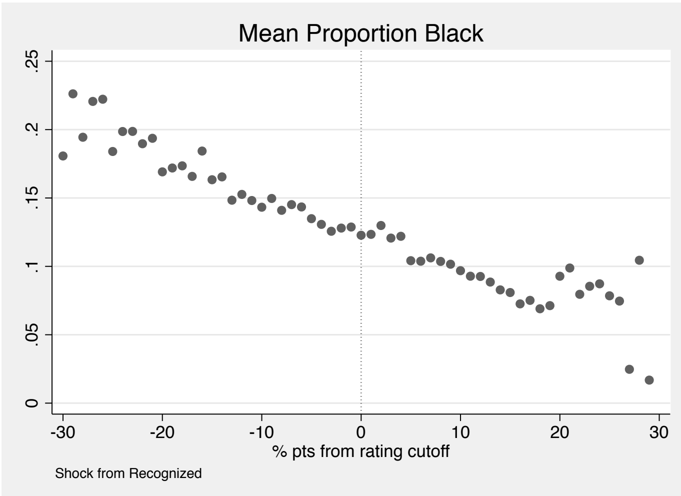
(d)



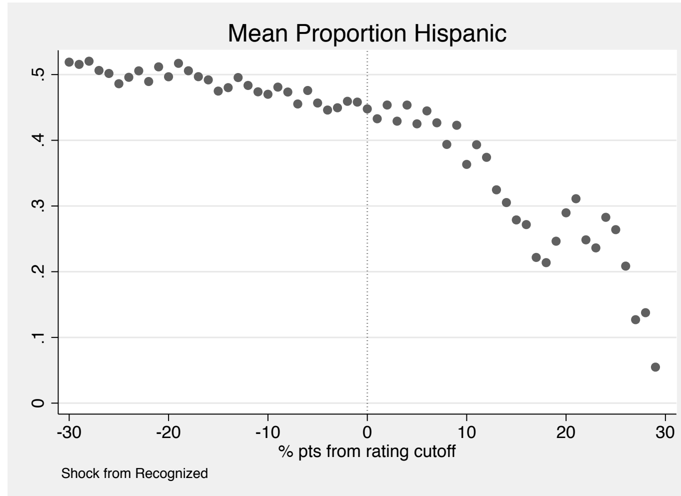
(e)



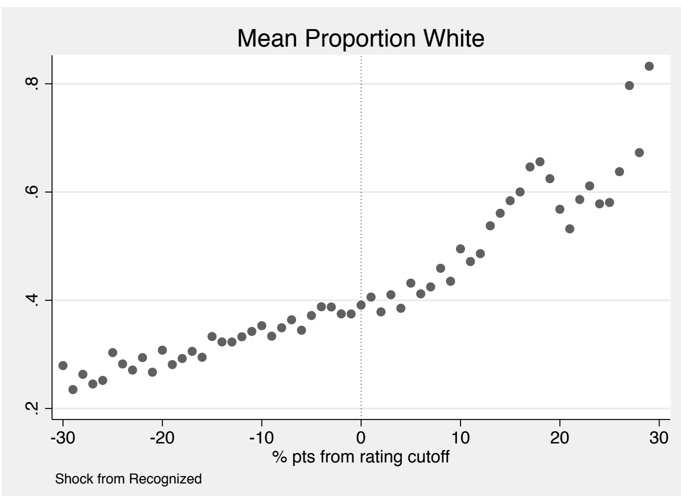
(f)


Mean Proportion Black
% pts from rating cutoff
Shock from Recognized

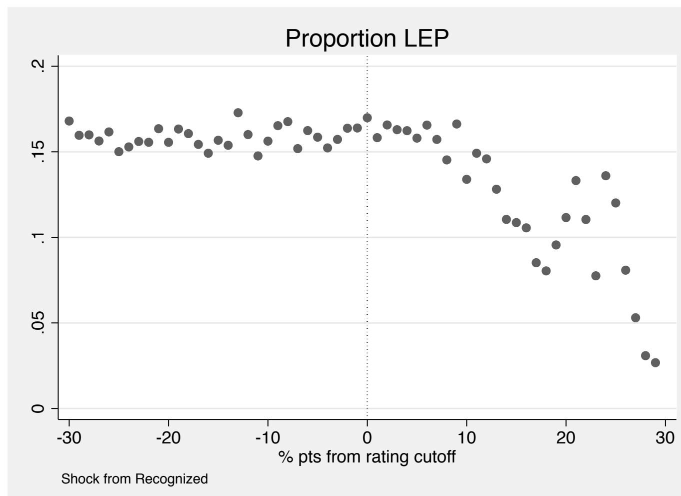
(g)


Mean Proportion Hispanic
% pts from rating cutoff
Shock from Recognized

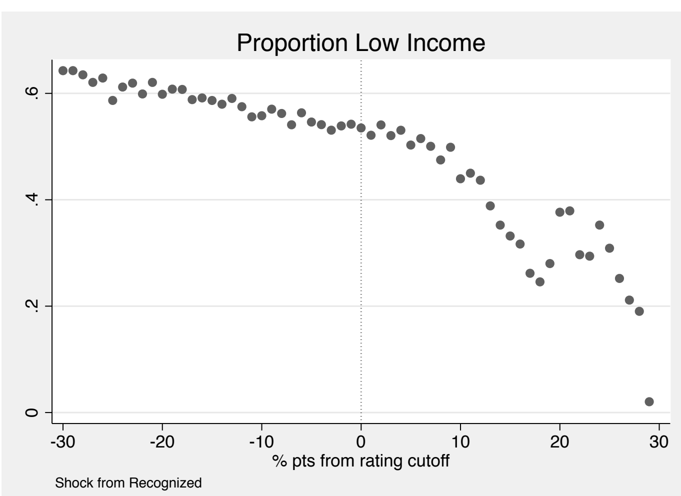
(h)


Mean Proportion White
% pts from rating cutoff
Shock from Recognized

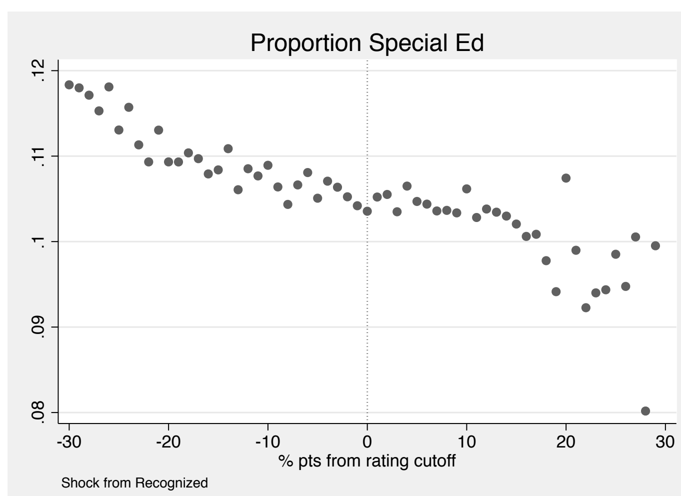
(i)


Proportion LEP
% pts from rating cutoff
Shock from Recognized

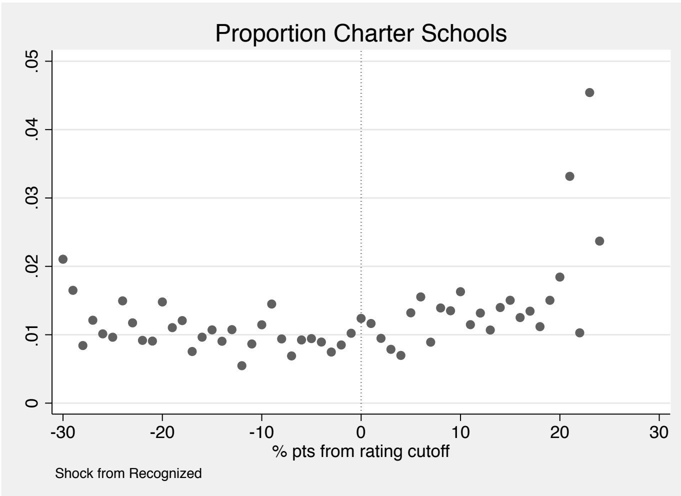
(j)


Proportion Low Income
% pts from rating cutoff
Shock from Recognized

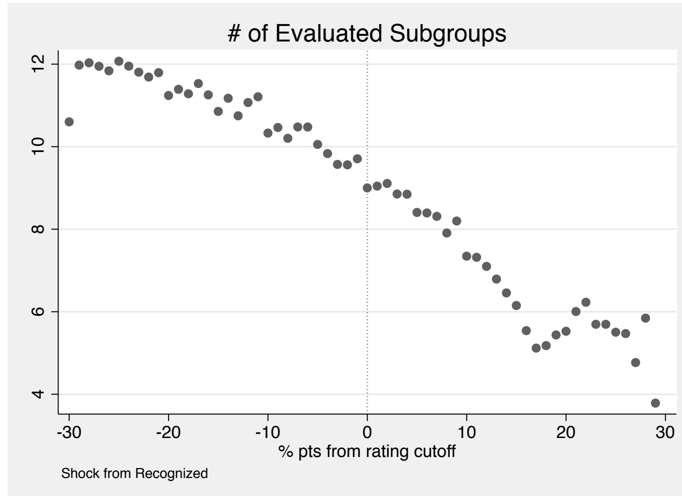
(k)


Proportion Special Ed
% pts from rating cutoff
Shock from Recognized

(l)

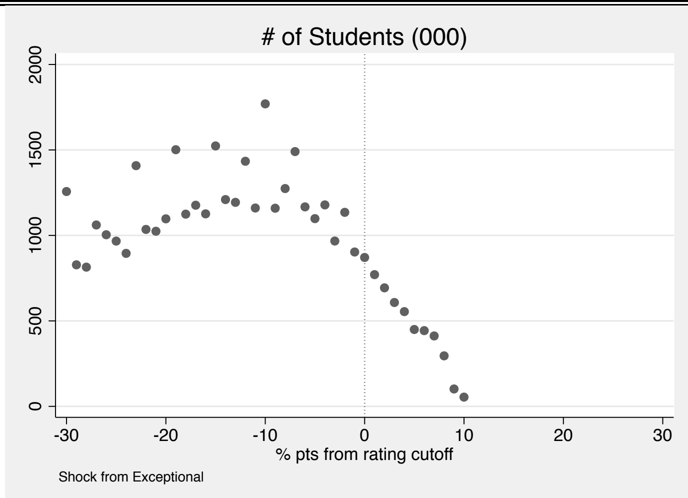

Proportion Charter Schools
.05
.04
.03
.02
.01
0
-30
-20
-10
0
10
20
30
% pts from rating cutoff
Shock from Recognized

(m)

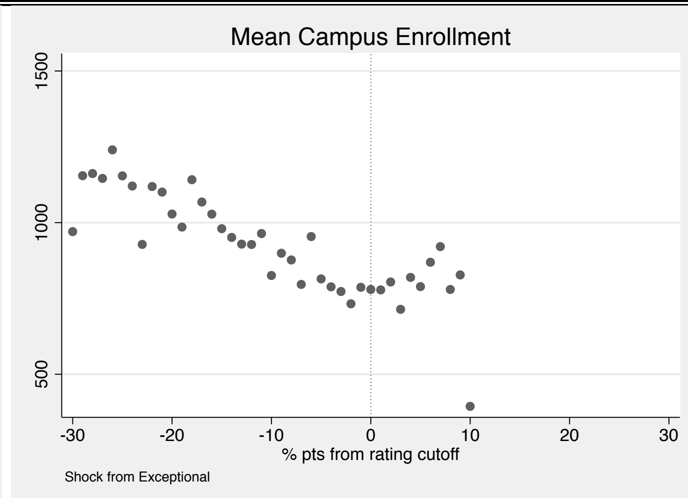

# of Evaluated Subgroups
12
10
8
6
4
-30
-20
-10
0
10
20
30
% pts from rating cutoff
Shock from Recognized

(n)

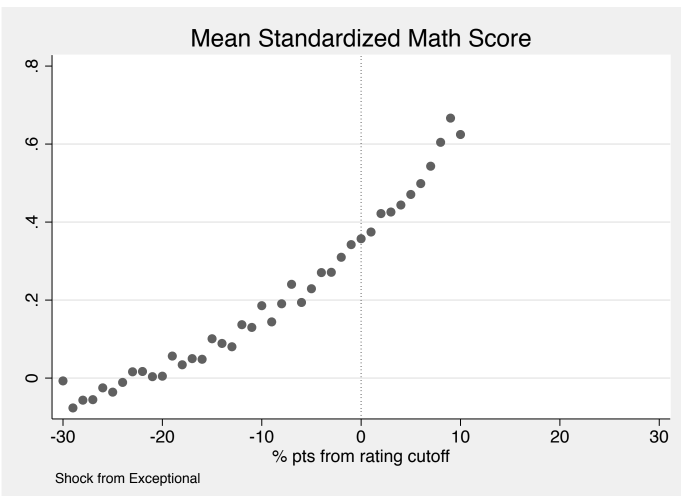
Figure 17. RD Assumptions: *Exemplary-Recognized* Pass-rate Boundary.



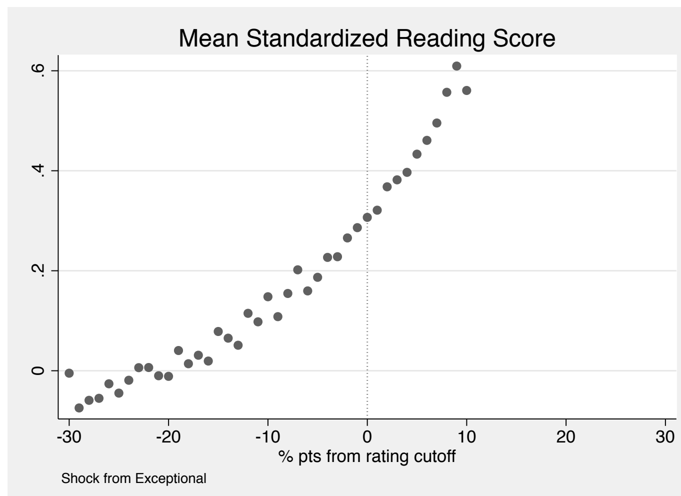
(a)



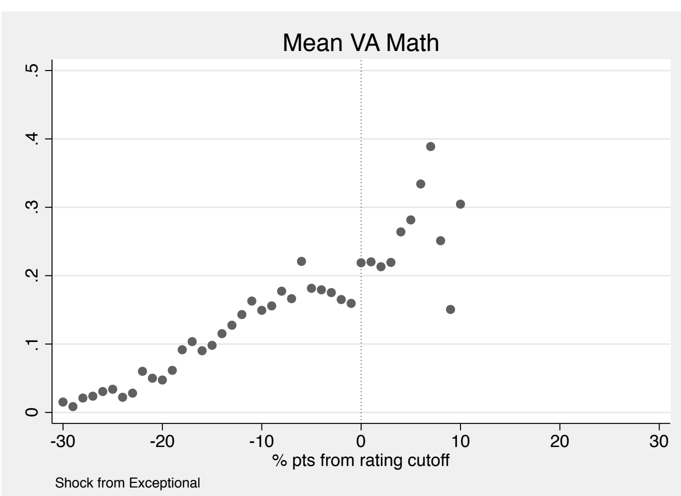
(b)



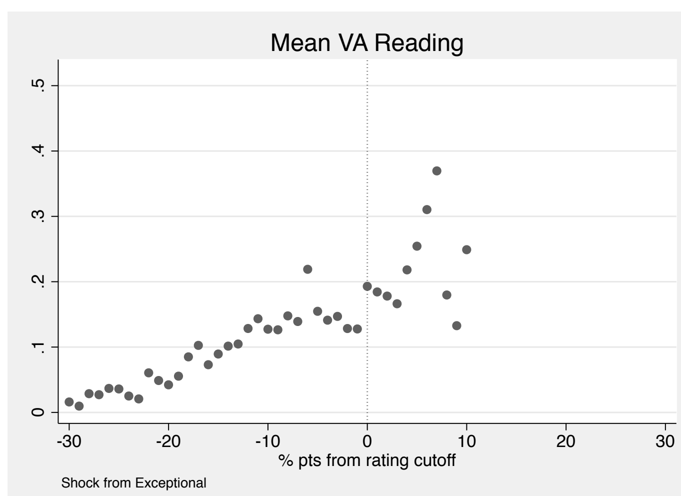
(c)



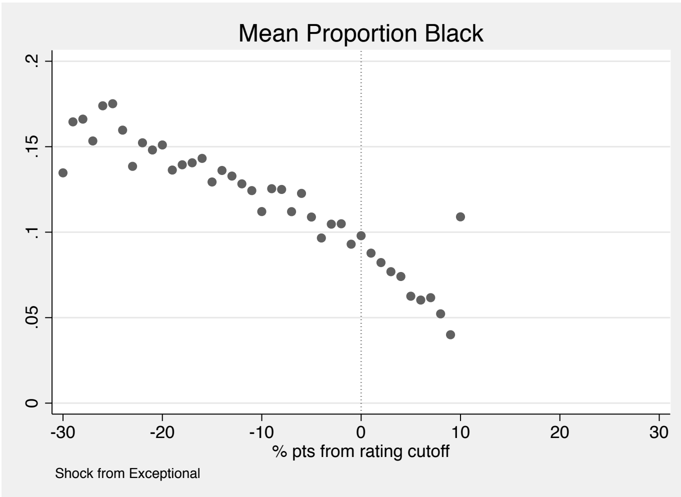
(d)



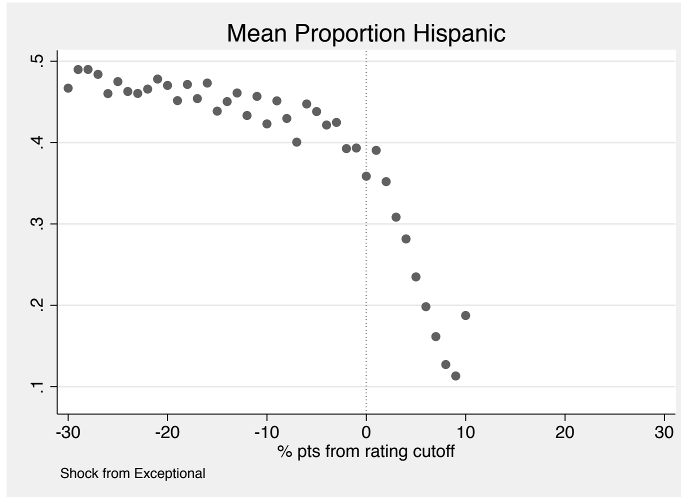
(e)



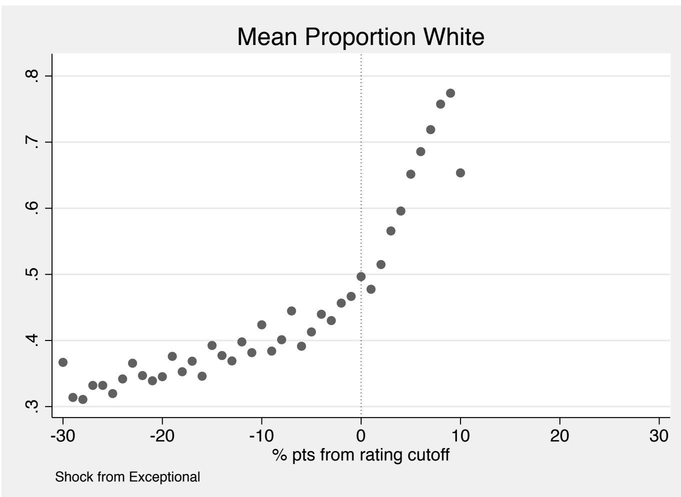
(f)


Mean Proportion Black
% pts from rating cutoff
Shock from Exceptional

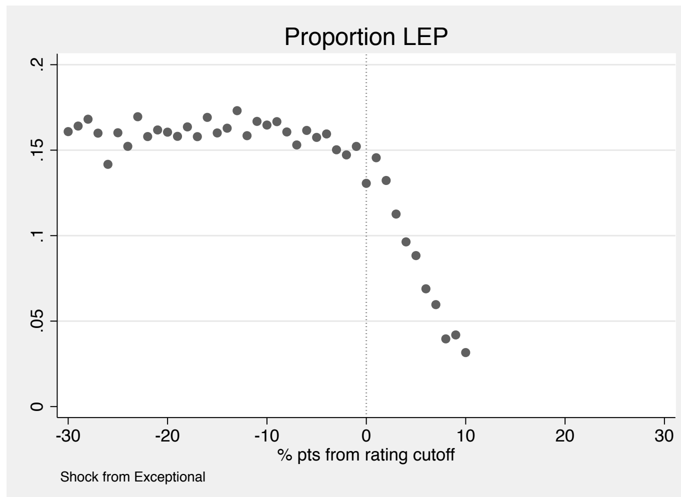
(g)


Mean Proportion Hispanic
% pts from rating cutoff
Shock from Exceptional

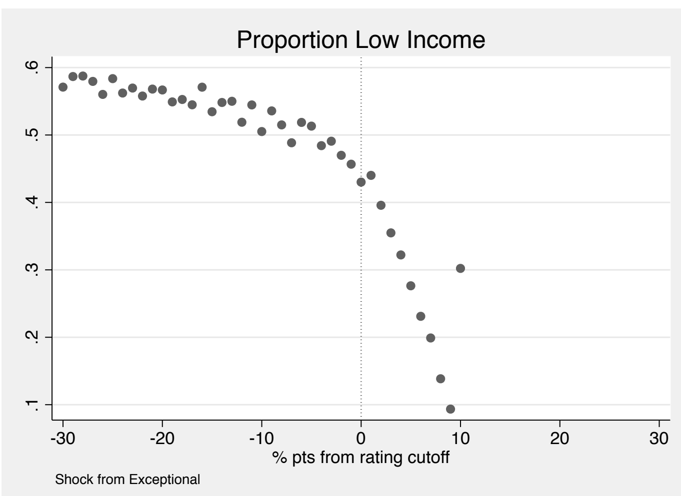
(h)


Mean Proportion White
% pts from rating cutoff
Shock from Exceptional

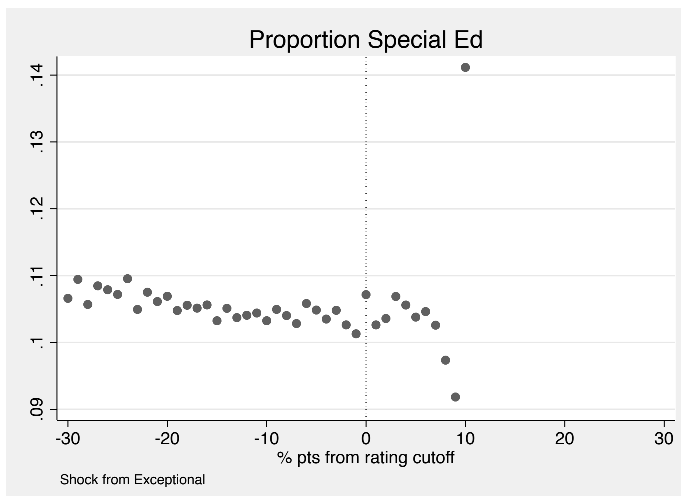
(i)


Proportion LEP
% pts from rating cutoff
Shock from Exceptional

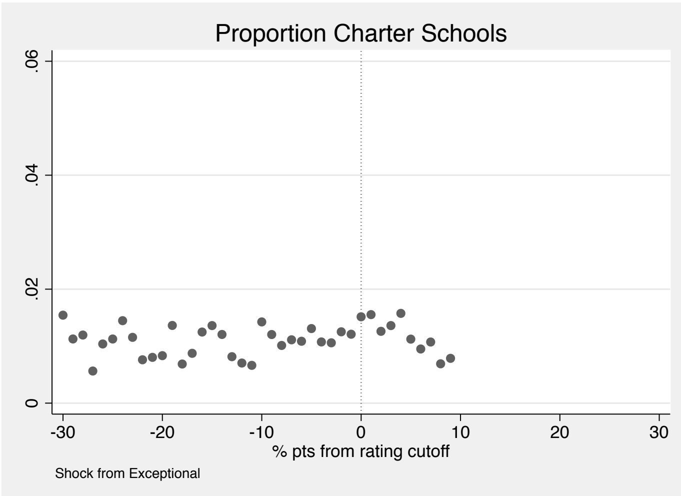
(j)


Proportion Low Income
% pts from rating cutoff
Shock from Exceptional

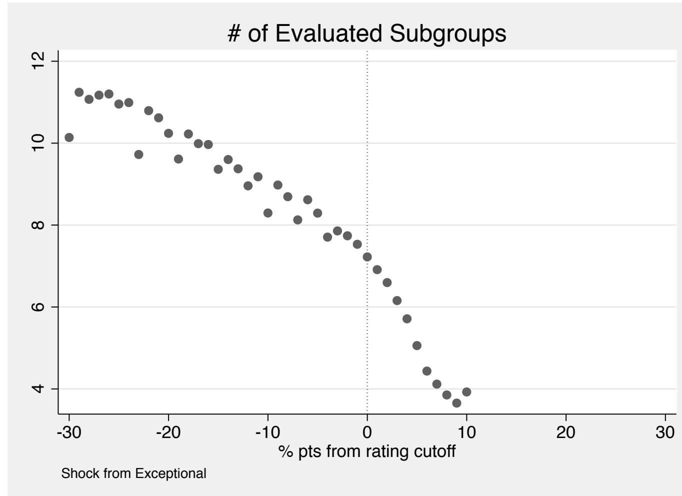
(k)


Proportion Special Ed
% pts from rating cutoff
Shock from Exceptional

(l)

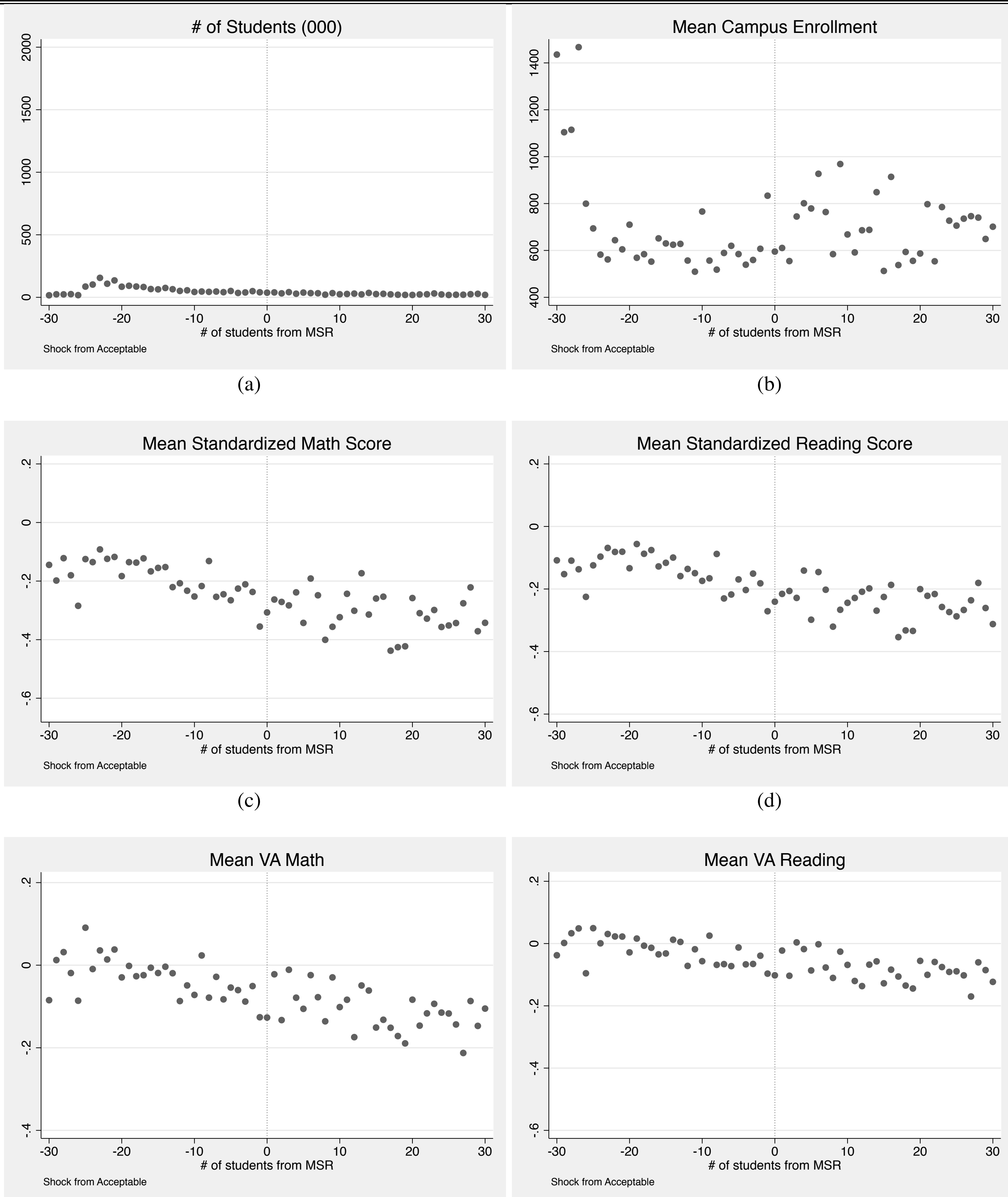

Proportion Charter Schools
0
.02
.04
.06
-30
-20
-10
0
10
20
30
% pts from rating cutoff
Shock from Exceptional

(m)


# of Evaluated Subgroups
4
6
8
10
12
-30
-20
-10
0
10
20
30
% pts from rating cutoff
Shock from Exceptional

(n)

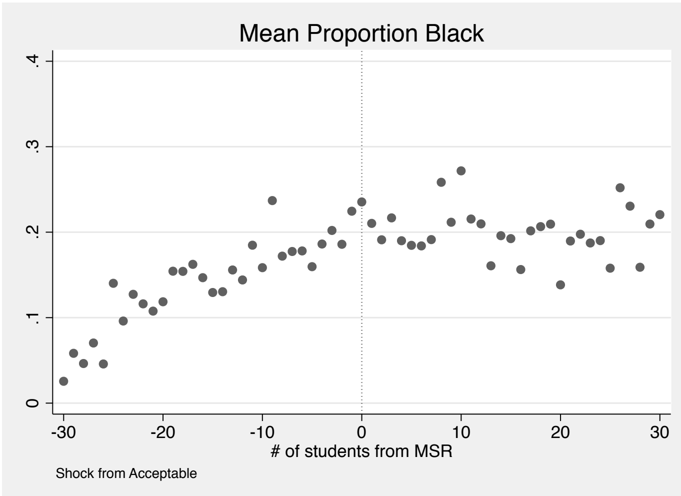
Figure 18. RD Assumptions: *Acceptable-Unacceptable* Group Size Boundary.

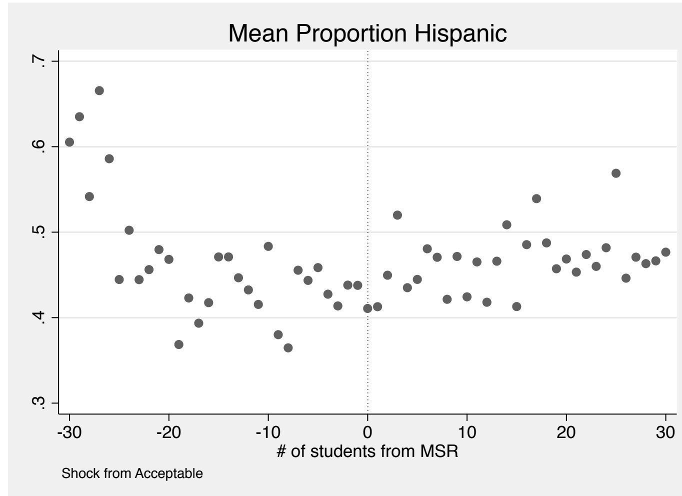


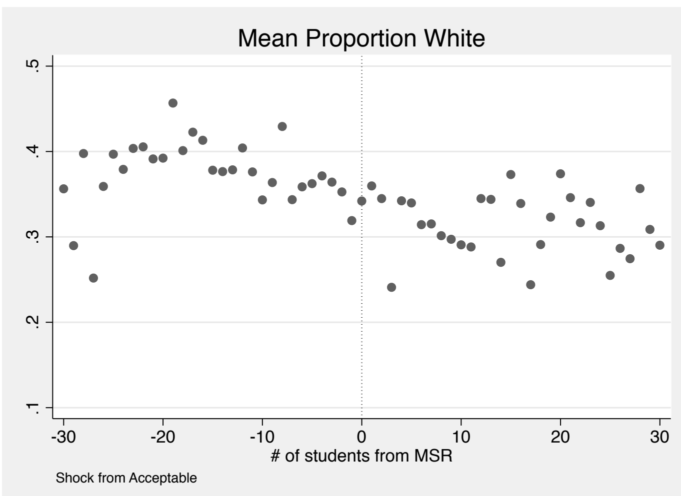
(a) (b)

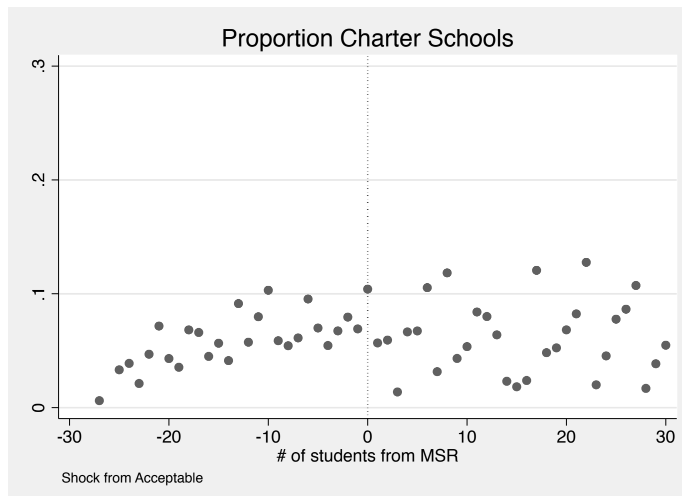
(c) (d)

(e) (f)


Mean Proportion Black
# of students from MSR
Shock from Acceptable
(g)


Mean Proportion Hispanic
# of students from MSR
Shock from Acceptable
(h)


Mean Proportion White
# of students from MSR
Shock from Acceptable
(i)


Proportion Charter Schools
# of students from MSR
Shock from Acceptable
(j)

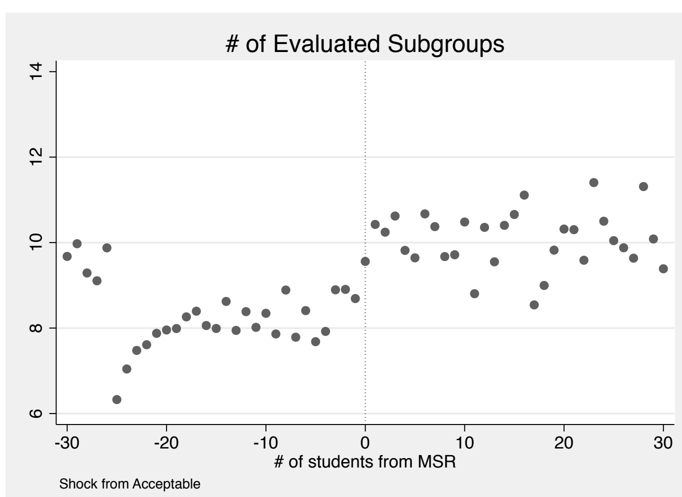
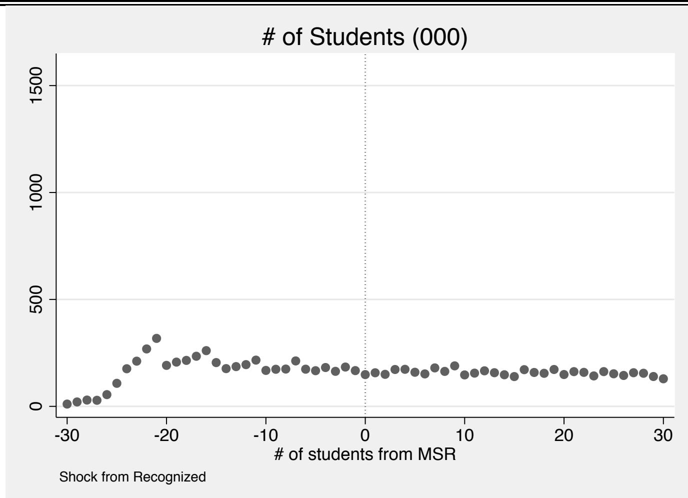
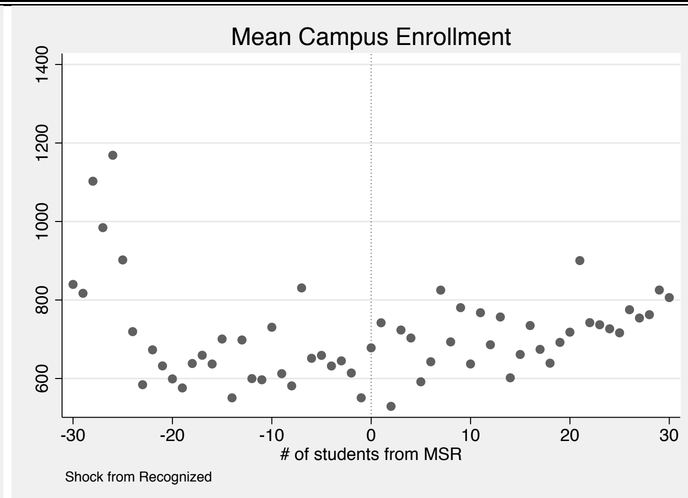

# of Evaluated Subgroups
# of students from MSR
Shock from Acceptable
(k)

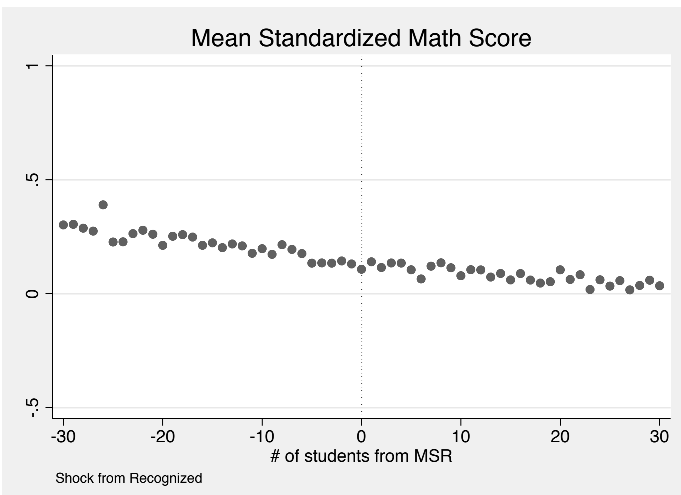
Figure D5. RD Assumptions: *Recognized-Acceptable* Group Size Boundary.



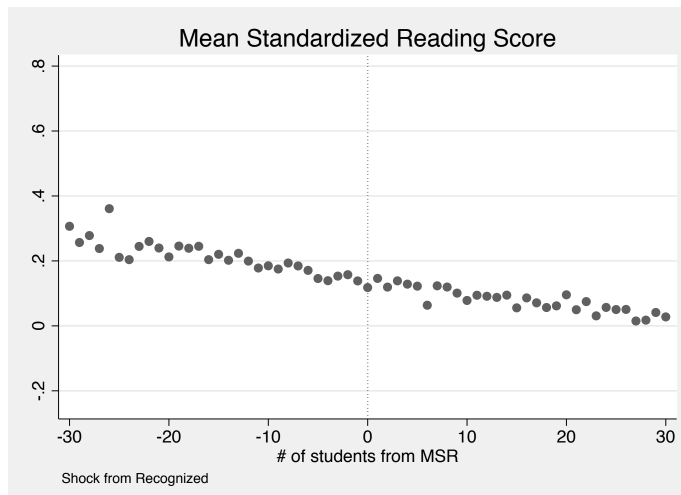
(a)



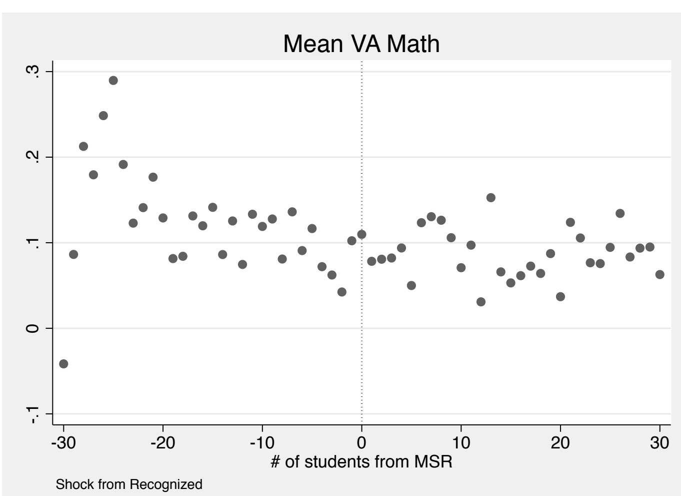
(b)



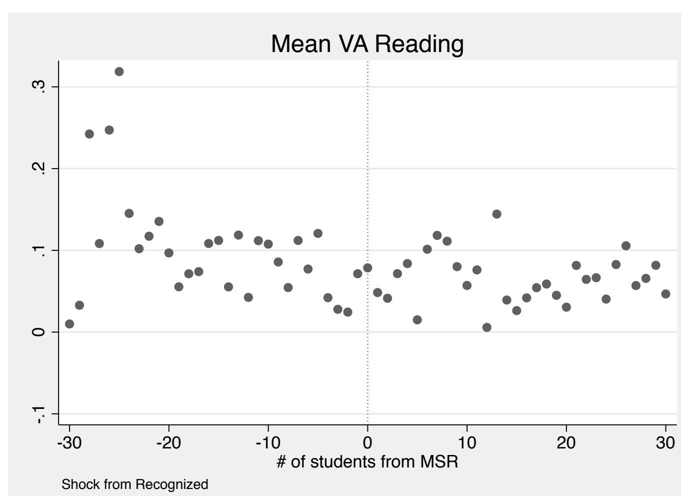
(c)



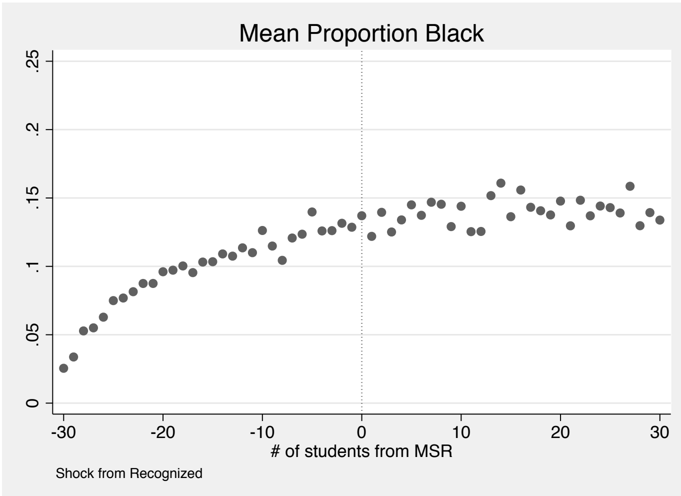
(d)



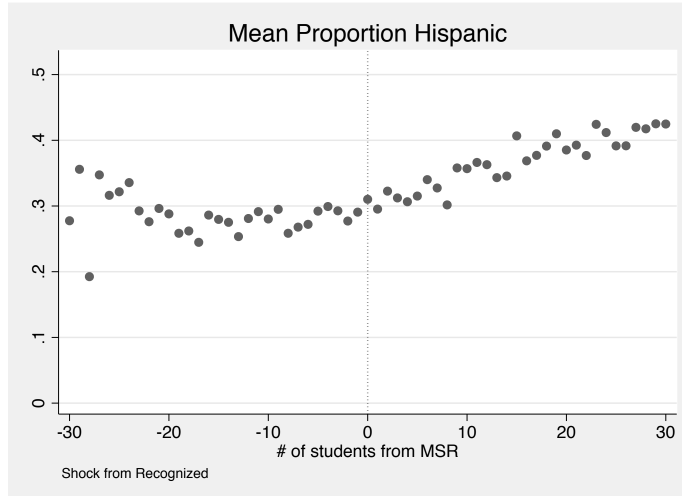
(e)



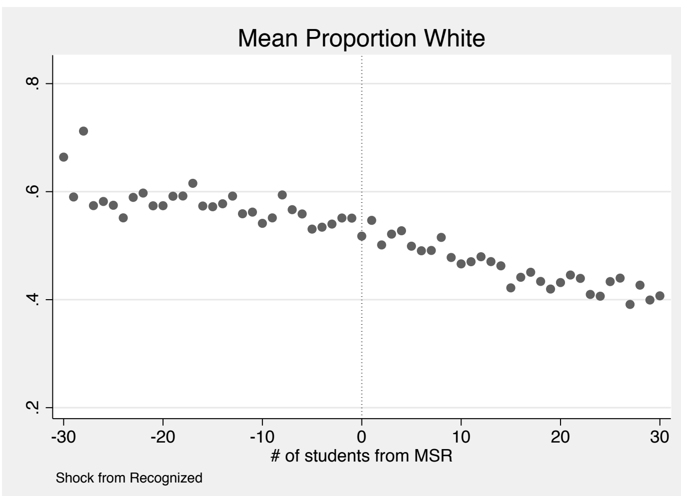
(f)


Mean Proportion Black
0
.05
.1
.15
.2
.25
-30
-20
-10
0
10
20
30
# of students from MSR
Shock from Recognized

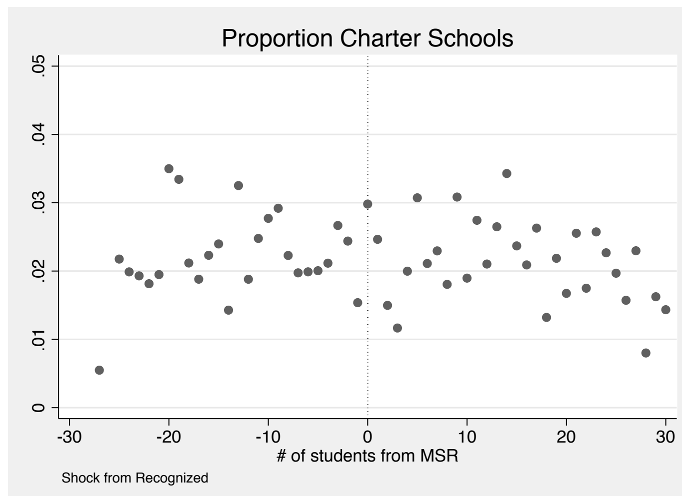
(g)


Mean Proportion Hispanic
0
.1
.2
.3
.4
.5
-30
-20
-10
0
10
20
30
# of students from MSR
Shock from Recognized

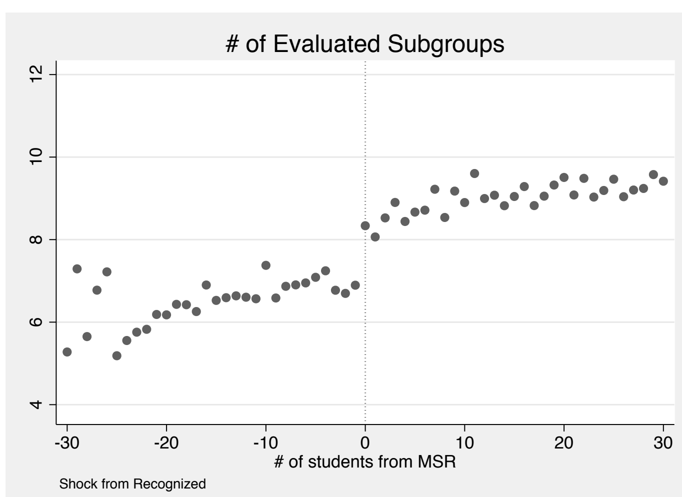
(h)


Mean Proportion White
.2
.4
.6
.8
-30
-20
-10
0
10
20
30
# of students from MSR
Shock from Recognized

(i)

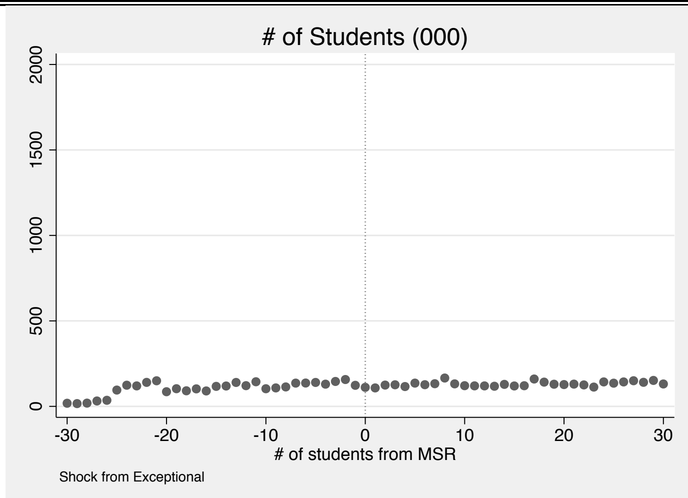

Proportion Charter Schools
0
.01
.02
.03
.04
.05
-30
-20
-10
0
10
20
30
# of students from MSR
Shock from Recognized

(j)

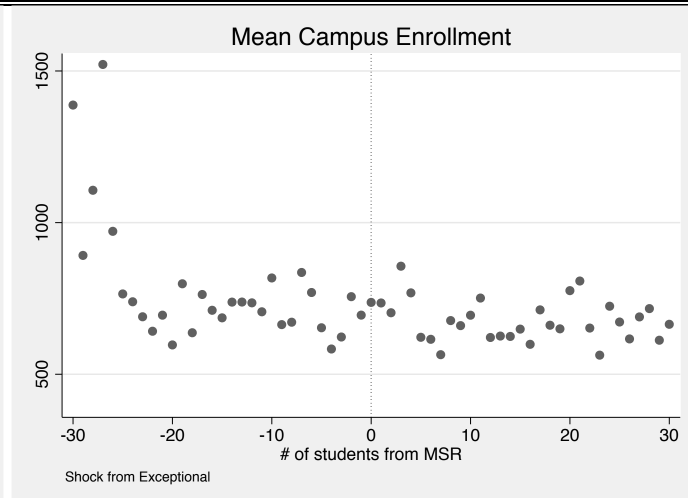

# of Evaluated Subgroups
4
6
8
10
12
-30
-20
-10
0
10
20
30
# of students from MSR
Shock from Recognized

(k)

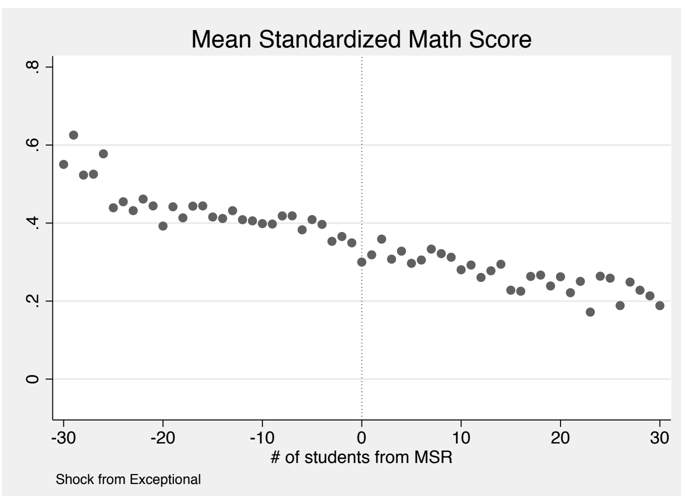
Figure 19. RD Assumptions: *Exemplary-Recognized* Group Size Boundary.



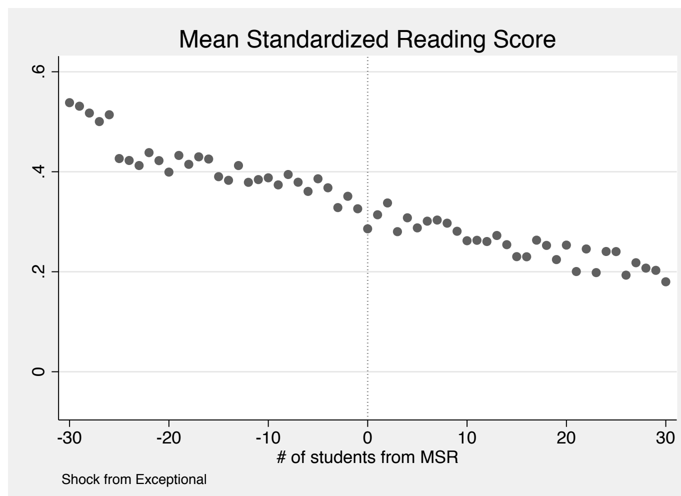
(a)



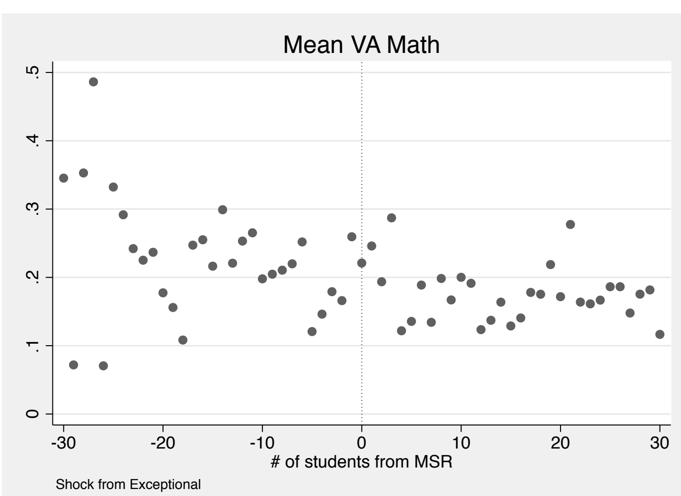
(b)



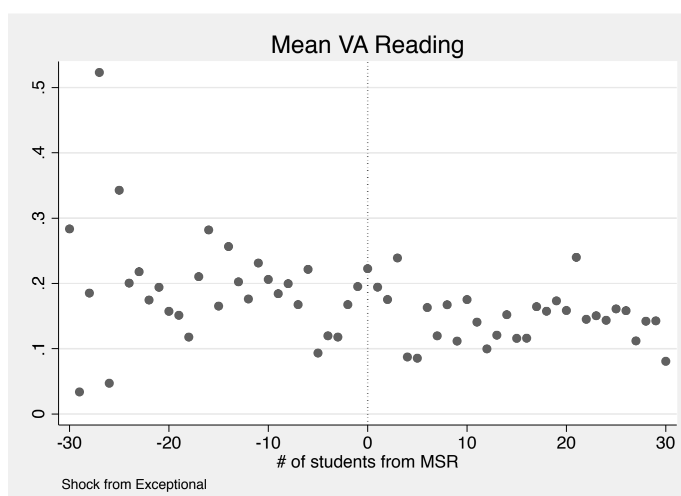
(c)



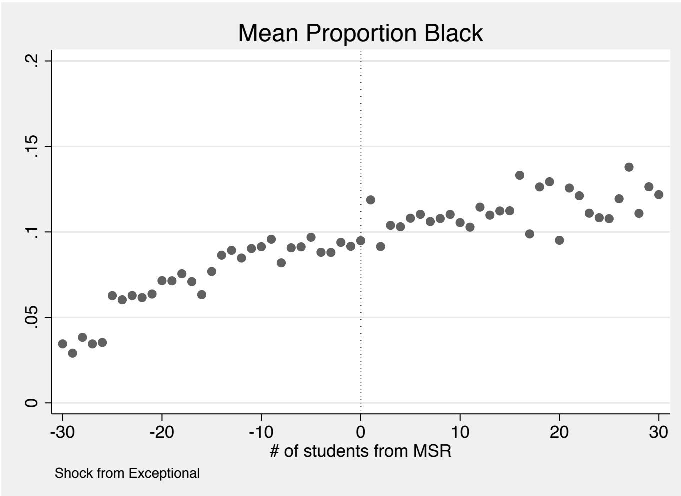
(d)



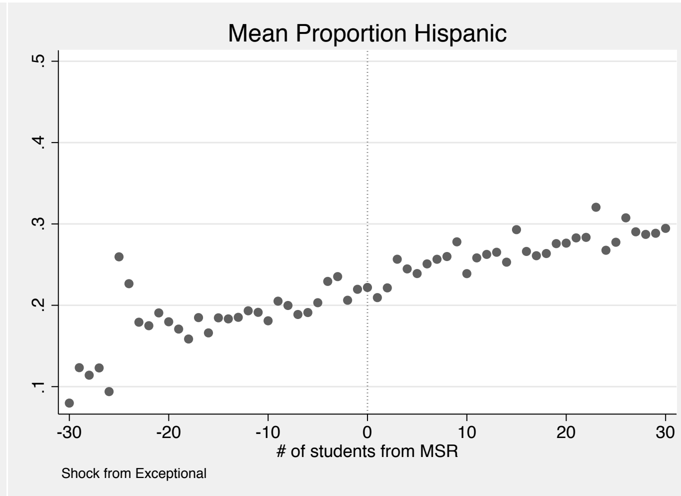
(e)



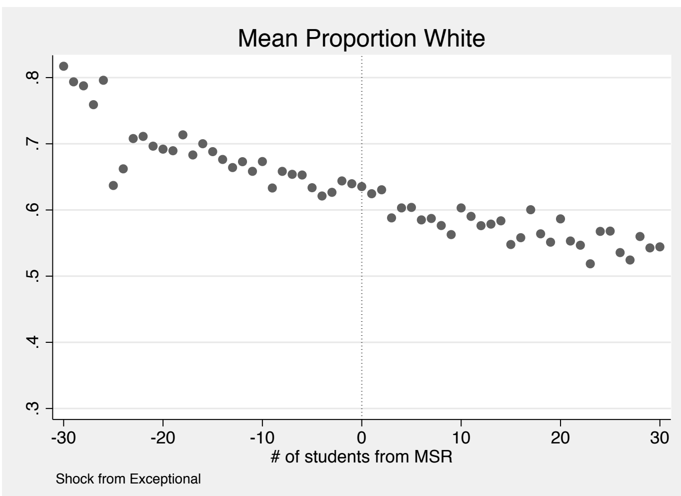
(f)


Mean Proportion Black
0
.05
.1
.15
.2
-30
-20
-10
0
10
20
30
# of students from MSR
Shock from Exceptional

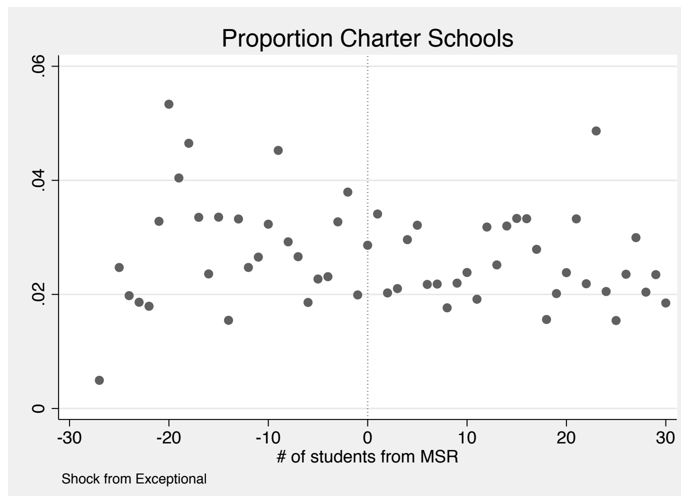
(g)


Mean Proportion Hispanic
.1
.2
.3
.4
.5
-30
-20
-10
0
10
20
30
# of students from MSR
Shock from Exceptional

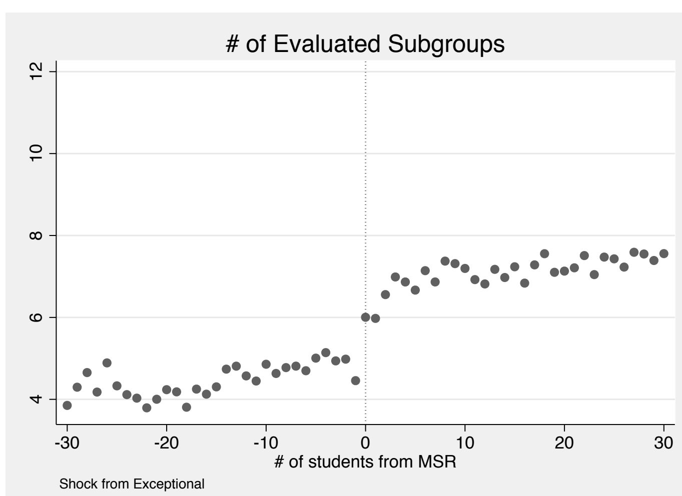
(h)


Mean Proportion White
.3
.4
.5
.6
.7
.8
-30
-20
-10
0
10
20
30
# of students from MSR
Shock from Exceptional

(i)


Proportion Charter Schools
0
.02
.04
.06
-30
-20
-10
0
10
20
30
# of students from MSR
Shock from Exceptional

(j)


# of Evaluated Subgroups
4
6
8
10
12
-30
-20
-10
0
10
20
30
# of students from MSR
Shock from Exceptional

(k)

APPENDIX C - Texas Accountability Rating System: Relevant Details

Exams Regimes

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; 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. 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.

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.

Nomenclature

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.

Minimum Size Requirement

The Minimum size requirement can be succinctly expressed as min(max(30,examinees/10),50) 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.

Other Factors

In addition to the performance on standardized tests, various other factors entered the rating formula over time. These include: the annual dropout rate amongst 7th and 8th graders; the annual dropout rate among 7th-12th graders; the four-year completion rate amongst 9th-12th graders; the State Developed Alternative Assessment (SDAA and SDAA II) for special education students between ; and English language learner (ELL) and commended performance rates In 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.

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 TPM allowed 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.

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. In 2009 and 2010 the TEA used the Texas Projection Measure (TPM) when computing student performance. (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. Details of the calculation of the TPM are beyond the scope of this paper. For more details, see the TEA's description at : http://www.tea.state.tx.us/index3.aspx?id=8351&menu_id=793.

Rewards & Interventions

Between 1992-2006 the Texas Successful Schools Award System (TSSAS) provided funds to schools using a value added or gains approach. Although I do not have data on TSSAS funding, these awards are based on value added and therefore are not likely to be discontinuous at the discontinuities used in this paper. Nevertheless, I outline the general scheme of the 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.

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."

TEA Guiding Principles

The TEA outlines eight guiding principles in the construction of the school accountability rating system. These principles are outlined in the annual accountability manual:

(1) STUDENT PERFORMANCE: The system is first and foremost designed to improve student performance.

(2) RECOGNITION OF DIVERSITY: The system is fair and recognizes diversity among campuses and students.

(3) SYSTEM STABILITY: The system is stable and provides a realistic, practical timeline for measurement, data collection, planning, staff development, and reporting;

(4) STATUTORY COMPLIANCE: The system is designed to comply with statutory requirements.

(5) APPROPRIATE CONSEQUENCES: The system sets reasonable standards for adequacy, identifies and publicly recognizes high levels of performance and performance improvement, and identifies campuses with inadequate performance and provides assistance.

(6) LOCAL PROGRAM FLEXIBILITY: The system allows for flexibility in the design of programs to meet the individual needs of students.

(7) LOCAL RESPONSIBILITY: The system relies on local school districts to develop and implement local accountability systems that complement the state system.

(8) PUBLIC'S RIGHT TO KNOW: The system supports the public's right to know levels of student performance in each school district and on each campus.

Table 21. Minimum pass-rate requirements (%) for each rating level by year and subject.

	Exemplary					Recognized					Acceptable				
	m	r	w	s	c	m	r	w	s	c	m	r	W	s	c
1994	90	90	90			65	65	65			25	25	25		
1995	90	90	90			70	70	70			25	25	25		
1996	90	90	90			70	70	70			30	30	30		
1997	90	90	90			75	75	75			35	35	35		
1998	90	90	90			80	80	80			40	40	40		
1999	90	90	90			80	80	80			45	45	45		
2000	90	90	90			80	80	80			50	50	50		
2001	90	90	90			80	80	80			50	50	50		
2002	90	90	90	90		80	80	80	80		55	55	55	50	
2003															
2004	90	90	90	90	90	70	70	70	70	70	35	50	50	50	25
2005	90	90	90	90	90	70	70	70	70	70	35	50	50	50	25
2006	90	90	90	90	90	70	70	70	70	70	40	60	60	60	35
2007	90	90	90	90	90	75	75	75	75	75	45	65	65	65	40
2008	90	90	90	90	90	75	75	75	75	75	50	70	65	65	45
2009	90	90	90	90	90	75	75	75	75	75	55	70	70	70	50
2010	90	90	90	90	90	80	80	80	80	80	60	70	70	70	55
2011	90	90	90	90	90	80	80	80	80	80	65	70	70	70	60

Notes: The headings m, r, w, s, and c stand for mathematics, reading/ela, writing, social studies and science, respectively.

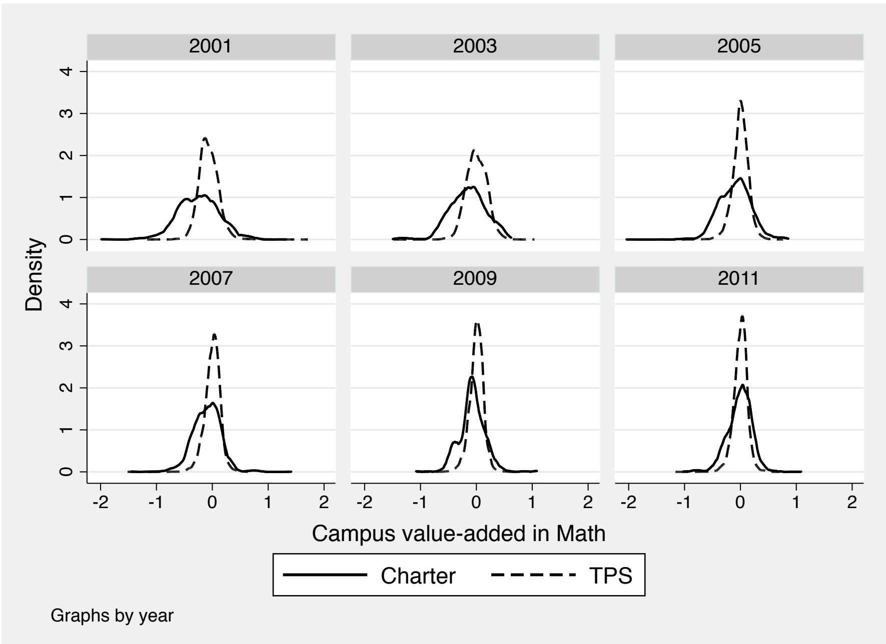
Table 22. Exceptions Provisions (Taken from 2006 Accountability Manual)

Number of Assessment Measures Evaluated	Maximum Number of Exceptions Allowed
1–5	0 exceptions
6 – 10	1 exception
11 – 15	2 exceptions
16 or more	3 exceptions

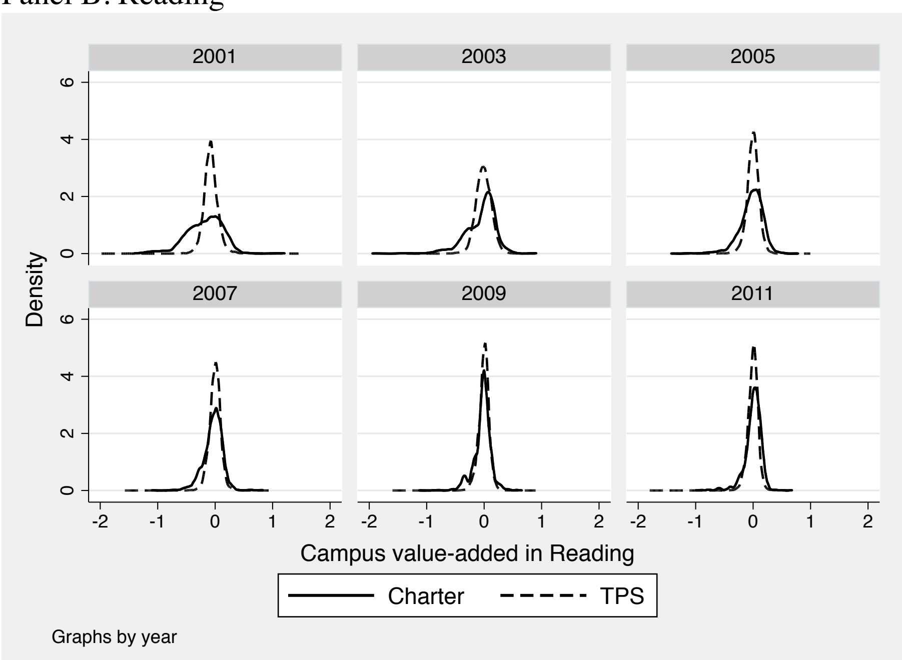
APPENDIX D – Distributions of Charter Quality

Figure 20. Distributions of School Quality by Year (Statewide estimates)

Panel A. Mathematics



Panel B. Reading



Notes: Distributions of residual quality after controlling for 1, 2, 3, 4, and 5+ years of operation. Kolmogorov – Smirnov tests performed for differences in distribution for each comparison year all yield p-values that indicate significant differences at the 1 percent level.

APPENDIX E – Descriptive profiles by receiving school rating

Table 23. Incoming student and sending school characteristics

Panel A. Incoming Student Characteristics

	Exemplary	Recognized	Acceptable	Unacceptable
Domographics				
White	0.542	0.427	0.335	0.189
	(0.002)	(0.001)	(0.001)	(0.005)
Black	0.092	0.121	0.173	0.317
	(0.001)	(0.000)	(0.001)	(0.006)
Hispanic	0.312	0.421	0.471	0.483
	(0.002)	(0.001)	(0.001)	(0.007)
Low Income	0.434	0.584	0.651	0.751
	(0.002)	(0.001)	(0.001)	(0.004)
Limited English	0.131	0.170	0.180	0.176
	(0.001)	(0.001)	(0.001)	(0.004)
Mean Cohort Size	549	592	707	616
	(2.968)	(2.620)	(3.133)	(12.45)
N	13650	26937	31633	1769
Achievement				
Math	0.110	-0.067	-0.196	-0.519
	(0.004)	(0.002)	(0.002)	(0.011)
Reading	0.106	-0.063	-0.181	-0.472
	(0.004)	(0.002)	(0.002)	(0.011)
N	12763	25426	30288	1700

Panel B. Sending School Characteristics

	Exemplary	Recognized	Acceptable	Unacceptable
Value Added				
Math	0.047	0.030	-0.004	-0.033
	(0.000)	(0.000)	(0.000)	(0.003)
Reading	0.027	0.007	-0.023	-0.039
	(0.000)	(0.000)	(0.000)	(0.003)
N	12,528	25,097	28,510	1,675
Mean Achievement				
Math	0.129	0.022	-0.085	-0.284
	(0.001)	(0.001)	(0.001)	(0.005)
Reading	0.123	0.016	-0.082	-0.263
	(0.001)	(0.001)	(0.001)	(0.005)
N	13,633	26,927	31,623	1,766
Proportion Charter	0.019	0.018	0.018	0.038
	(0.000)	(0.000)	(0.000)	(0.001)
N	13,633	26,927	31,623	1,766

Notes: Panel (a) summarizes characteristics of incoming students. Panel (b) summarizes the characteristics of the sending schools. Standard errors in parentheses.

VITA

NAME:	Patrick Lawrence Baude
EDUCATION:	B.S.E., Electrical Engineering, University of Pennsylvania, Philadelphia, PA, 2003.
	M.S., Economics, State University of New York at Buffalo, Buffalo, NY, 2010
TEACHING:	Department of Economics, University of Illinois at Chicago; Econometrics (Undergraduate), 2014-2015.
HONORS:	James Doti Award, Department of Economics, University of Illinois at Chicago, 2014
	Mitchell Krask Fellowship, Department of Economics, University of Illinois at Chicago, 2010
PRESENTATIONS:	Association of Education, Finance, and Policy Annual Conference, San Antonio, TX, March 15, 2014
	CERGE-EI Skills Workshop, Prague, Czech Republic, September 16, 2014
	University of Illinois at Chicago Economics Club, Chicago, IL, October 29, 2014
	Association of Public Policy Analysis and Management Fall Meeting, Albuquerque, NM, November 6, 2014
	Department of Economics Active Research Lunch, University of Illinois at Chicago, Chicago, IL, November 4, 2015
PROFESSIONAL EXPERIENCE:	Sudbury Neutrino Observatory, Philadelphia, 2001
	Federal Home Loan Mortgage Corporation, McLean, VA, 2002, 2003
	Citigroup Global Markets, Fixed Income Research, New York, NY, London, UK, Buffalo, NY, 2003-2010