School Accountability and Teacher Labor Markets

Andrew Smith[*]
University of Wisconsin–Madison

September 14, 2024

Abstract

I study the impact of school and district accountability pressures on elementary school teacher labor markets and student performance in Wisconsin. I find a strong inverse relationship between accountability scores and teacher turnover. Using a regression discontinuity design, I find limited evidence that teacher turnover discontinuously changes at any cutoff, even when conditioning on value-added. Moreover, changes in school level value-added and student test performance appear unaffected by being on either side of a cutoff.

1 Introduction

School accountability has been at the forefront of education reform in the United States for the past three decades. Developed in response to reports like *A Nation at Risk*, and institutionalized through federal policies such as No Child Left Behind (NCLB) and the Every Student Succeeds Act (ESSA), accountability systems evaluate educators primarily through their students' performance on standardized tests. These policies have concerned many policymakers, as they often allow for strategic behavior by school administrators and teachers.

Absent incentives, teachers largely prefer the lower stress environment that comes with teaching at a high-performing school (Jones et al., 1999; Kirtley, 2012). However, high-quality teachers are wanted at low-performing schools, as they are most capable of improving long-run outcomes for the students at these schools (Chetty et al., 2014). Moreover, school accountability policies, designed to improve student performance at all levels, have mixed outcomes for students and teachers. Previous literature finds that school accountability raises educational performance on

[*]I would like to thank Matt Wiswall, Chao Fu, Jeff Smith, Martin O'Connell, and Jesse Gregory for their invaluable advising and patient support. I would also like to thank Carl B. Frederick of the WDPI for his incredible help in accessing and understanding the data.

low-stakes tests in math and reading (see Deming and Figlio, 2016 for a review of this literature). Other literature has shown that accountability hurts low-performing schools by increasing teacher turnover (Clotfelter et al., 2004; Feng et al., 2010; Gjefsen and Gunnes, 2016).[1]

My goal is to understand how accountability scores affects teachers, students, and schools at large using a regression discontinuity design. Accountability scores correspond to school/district "ratings", and being at or near a rating cutoff could induce discontinuities if consequences are present. Specifically for teachers, I explore if teachers sort based on accountability grades, and whether accountability grades induce differing levels of teacher exit rates. For students, I test whether accountability grades produce improvements in student standardized test performance. For schools, I evaluate the change in a school's average value-added, conditional on accountability grades.

This paper contributes to the understanding of accountability policy, and its effects on teacher labor markets, students, and schools (Clotfelter et al., 2004; Dizon-Ross, 2018; Feng et al., 2010; Gjefsen and Gunnes, 2016). Because Wisconsin's Act 10 allows for district to flexibly pay individual teachers, the Wisconsin setting allows for a more comprehensive view of this policy, as districts can recruit talent through salaries.[2] My design-based approach shows that teacher turnover and accountability scores are inversely related. Moreover, my regression discontinuity design estimates show limited evidence that schools and districts are responsive to being close to accountability score cutoffs.

The rest of the paper proceeds as follows: Section 2 examines the institutional setting in Wisconsin. Section 3 describes the data, measurement, and some preliminary evidence. Section 4 presents the design-based strategy, and Section 5 provides results. Section 6 concludes.

2 Institutional Setting

2.1 Accountability

In anticipation of the ESSA, the state of Wisconsin implemented Statute 115.385, which required the Wisconsin Department of Public Instruction (WDPI) provide annual School Report Cards to all public schools beginning with the 2011-12 school year. In the following academic year, school districts began receiving annual District Report Cards. The purpose of these report cards is to provide a rating system that effectively measures school and district relative performance, so that

[1] Dizon-Ross (2018) is at odds with this literature, showing that accountability in New York City actually decreased turnover for low-performing schools. She speculates that this is due to the high stakes nature of accountability in her setting.

[2] See Biasi (2021) and Biasi et al. (2021) for the effect of district-level wage strategies on teacher labor market decisions.

the state can effectively identify low-performing schools and districts in need of improvement. Annual report cards are publicly shared with educators, parents, and policymakers in an effort to keep schools and districts accountable.

When developing the report cards, WDPI focuses on four priority areas: academic achievement, year-on-year growth in academic achievement, outcomes for low-performing students, and graduation rates (WDPI, 2021d). For school accountability scores, these metrics are drawn from the performance of their full-time students. For district accountability scores, students from all schools are pooled together, as if the district is "one big school" (WDPI, 2022a). The overall accountability scores (which range from 0 – 100) are calculated by taking a weighted average of the scores from the four priority areas, resulting in the ratings and score ranges provided in Table 1. The weights for these priority areas vary widely by school and district, as they are a function of student demographics, school type, and data availability.[3] An example of the front page of a school report card can be seen in Figure A.1.[4]

Table 1: Accountability Rating Score Ranges

Accountability Rating	Accountability Score Range
Significantly Exceeds Expectations	83 – 100
Exceeds Expectations	73 – 82.9
Meets Expectations	63 – 72.9
Meets Few Expectations	53 – 62.9
Fails to Meet Expectations	0 – 52.9

During my sample period, there is some variation in the public release of school and district report cards. Originally, Statute 115.385 required WDPI to publicly release accountability grades no later than September 30 every year. In April 2018, this statute was amended to allow WDPI to release accountability grades by November 30. Public release notwithstanding, the process and timing for producing these grades are based on the following steps (WDPI, 2022b):

1. Students take standarized tests during academic school year t.[5]

2. Standardized test scores are publicly released either in the late Spring or subsequent Summer of academic year t.

[3] Schools and districts are able to determine their accountability weights ex-ante by entering relevant data into the state's weighting calculator, which is readily accessible at https://oea-dpi.shinyapps.io/overall_weighting_calculator/.

[4] To see the full report card for this example school, visit https://www.mcfarland.k12.wi.us/district/prosp-schl-perf.cfm.

[5] Exam timing varied within my sample. For academic years ending in 2011-2014, the Wisconsin Knowledge and Concepts Examination (WKCE) was administered to students in grades 3 through 8 in the Fall. In 2015, the Badger Exam tested students in grades 3 through 8 in the Spring. From 2016-19, the Forward Exam replaced the Badger Exam, but the exam continued to be administered to students in grades 3-8 in the Spring.

3. Either prior to school year $t+1$, or in the early Fall of school year $t+1$, WDPI releases preliminary report cards to schools and districts only. Schools and districts are then allowed to review their results, and submit any inquiries if they believe there are errors.

4. After a review of all inquiries, WDPI publicly releases final report cards in the $t+1$ academic school year (either by September 30 or November 30, as previously mentioned).

A school or district's rating can come with consequences. If a school in a disadvantaged area scores exceptionally on the closing the gaps priority area, WDPI may bestow the school a "School of Recognition" award. If a school is deemed academically superior, or exhibits progress in closing achievement gaps among student subgroups, the state superintendent may nominate the school for the U.S. Department of Education's "Blue Ribbon", a prestigious national distinction for exemplary schools.

If a school or district repeatedly receives low accountability scores, the state can deem the school or district to be "in need of improvement".[6] Once improvement status is realized, Wisconsin Statute 118.42 allows the state superintendent to impose interventions that could include: required professional development, administrative layoffs, staff layoffs, charter restructuring, or school closure.

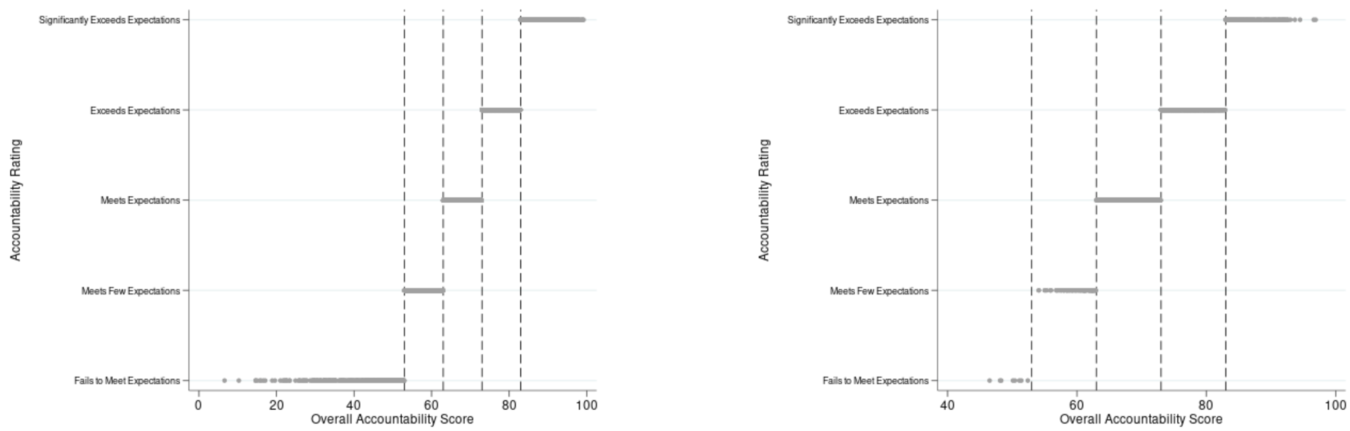
The ratings, embodied in a 0 – 100 numeric score, have strict cutoffs that have remained constant during the 2011-12 through 2013-14 and 2015-16 through 2018-19 school years.[7] Given the timing of these grades, the grade thresholds (absent manipulation) make Wisconsin's school and district accountability scores in academic year t a credible running variable in a sharp regression discontinuity design with multiple cutoffs in year $t+1$. In other words, I match teacher employment decisions to the lagged accountability scores of their current employer. Figure 1 illustrates the "sharpness" of these running variables. I discuss the validity of these running variables in Section 4.

2.2 Act 10

In line with the start of Wisconsin's new accountability regime, the state legislature also passed Act 10, which eliminated collective bargaining for general municipal employees. Within WDPI, this prevented the forced negotiation between teachers' unions and school districts, and gave districts full autonomy to negotiate directly with individual teachers. As highlighted by Biasi (2021),

[6] Improvement status is defined by Wisconsin PI 43, which is only realized after (1) repeatedly low scores are realized, (2) WDPI files a school improvement fund application to the U.S. Department of Education (DOE), and (3) the DOE approves the application. To see a list of schools and districts that are currently receiving intervention, visit https://dpi.wi.gov/title-i/1003g-school-improvement-grants.

[7] Wisconsin's accountability plan was suspended for the 2014-15 school year, as this policy was pending reapproval from the U.S. Department of Education.



(a) School Accountability Scores and Ratings

(b) District Accountability Scores and Ratings

Figure 1: Accountability Scores as Running Variables

districts primarily responded in two ways: half maintained a "seniority pay" scheme, which compensates teachers based on their experience, and the other half followed a "flexible pay" scheme, where teachers were compensated for both their experience and quality. Biasi (2021) shows that flexible pay raised salaries for high-quality teachers, and that high-quality teachers also moved from seniority pay to flexible pay districts.

There are a few ways the new accountability regime and Act 10 could interact. Given the flexibility Act 10 provides, schools and districts have the freedom to improve their accountability grades by attracting high-quality teachers through wages. Although this is not explicity addressed in my design-based approach, I do test whether being near a accountability rating cutoff affects the change in a school's teacher value-added. Moreover, districts could respond to accountability grades by updating their payment schemes from seniority pay to flexible pay. This potential strategy is not explored in this work, as there is no readily available panel data that records a district's payment schemes over time.

3 Data

3.1 Data Description

The first set of data I use comes from the *Public All Staff Report*, which is an administrative data set containing information on the universe of Wisconsin employees in the Wisconsin K-12 public school system (WDPI, 2021a,c). These data come from publicly available files downloaded from WDPI's website. These data include personal and demographic information, education, years of teaching experience, total salary, grades served, subjects taught, full-time equivalency (FTE) units, and school and district identifiers. These data were downloaded for the 2011-12 through 2018-19 school years.

The second set of data I have access to are the *Accountability Report Cards* (WDPI, 2021e). These data also come from publicly available files downloaded from WDPI's website. For schools, the data are available for the 2011-12 through 2013-14 and 2015-16 through 2018-19 school years. Outside of district accountability beginning in 2012-13, district data is similarly accessible. Data is not available for the 2014-15 school year, as Wisconsin was awaiting reapproval of their accountability policies from the U.S. Department of Education. These data include accountability scores, accountability ratings, the weights for the four accountability priority areas, the metrics used to compute accountability scores, school and district characteristics, and school and district identifiers.

The third data source I use comes from the *Wisconsin Information System for Education* (WISE). These data are confidential and were provided by the WDPI (WDPI, 2021b). This confidential record contains student-level demographics, attendance, discipline, test scores, and teacher-student linkages. WISE also provides annual measures of grade level value-added, which measures the value-added a "team" of teachers has on an entire grade of students.[8] These data are available for the 2009-10 through 2018-19 school years, except for teacher-student linkages, which WDPI began recording in the 2018-19 school year.[9]

The sample I construct focuses only on elementary school teachers that have an assignment with at least 0.5 full-time equivalent units. Moreover, I require that these teachers are hired by a public school district.[10] Given that the variation I am interested in exploring is a teacher's $t+1$ employment given her time t school's accountability grade, I build a panel of teacher employment decisions from the *Public All Staff Report* and match it to the lags of accountability scores from the *Accountability Report Cards* during the 2012-13 to 2018-19 school years. I develop the aforementioned outcome variable of interest, and match this panel with the rich school and district characteristics provided by WISE. Since the accountability system was suspended for the 2014-15 school year, I fail to have a lag for the 2015-16 school year, so employment decisions for the 2015-16 school year are currently ignored. When performing analysis that conditions on value-added, I restrict my sample to only include teachers who have estimates of value-added.[11]

[8] To see how Wisconsin typically estimates grade value added, refer to the following techical report: https://dpi.wi.gov/sites/default/files/imce/accountability/pdf/WI_DPI_School_VA_Technical_Report_2019.pdf

[9] For the majority of my data, teachers are only linked to the school grades they taught.

[10] This condition effectively excludes teachers who are subcontracted.

[11] I construct these measures of value-added. My methods of construction can be seen in Section 3.3.1.

3.2 Summary Statistics and Preliminary Evidence

Summary statistics for students, conditional on both school and district accountability ratings, can be seen in Table 2 and Table 3, respectively. Schools and districts with the lowest accountability ratings have, on average, larger student enrollments, and more demographic diversity. These schools and districts also have larger shares of students with limited English proficiency (LEP), special education status (SPED), and free or reduced price lunch status (FRPL). As expected, standardized test scores are increasing in accountability rating, but surprisingly, the average standardized test scores for the lowest rated schools are more than two standard deviations below the sample average. Accountability scores also are persisitent. Figure A.2 presents a binned scatterplots of school and district scores and their second lag. These figures show that these (binned) ordered pairs tightly hug the 45-degree line.

Table 2: School Summary Statistics, School Accountability

	Total	Fails to Meet Expectations	Meets Few Expectations	Meets Expectations	Exceeds Expectations	Sig. Exceeds Expectations
School Enrollment	377.4	412.9	407.9	368.7	376.3	374.1
	(169.4)	(150.7)	(167.6)	(168.0)	(168.2)	(177.4)
Asian Students	0.0356	0.0262	0.0420	0.0355	0.0341	0.0380
	(0.0565)	(0.0488)	(0.0634)	(0.0621)	(0.0511)	(0.0538)
Black Students	0.0907	0.694	0.332	0.0614	0.0371	0.0281
	(0.200)	(0.340)	(0.334)	(0.121)	(0.0755)	(0.0615)
Hispanic Students	0.114	0.120	0.236	0.125	0.0939	0.0732
	(0.145)	(0.189)	(0.241)	(0.149)	(0.113)	(0.0727)
LEP	0.0667	0.0653	0.140	0.0756	0.0542	0.0377
	(0.101)	(0.132)	(0.157)	(0.105)	(0.0800)	(0.0569)
Student Repeater	0.00403	0.00740	0.00440	0.00415	0.00297	0.00568
	(0.0340)	(0.0112)	(0.00825)	(0.0337)	(0.0236)	(0.0604)
FRPL	0.458	0.865	0.766	0.511	0.385	0.266
	(0.234)	(0.145)	(0.178)	(0.182)	(0.184)	(0.178)
SPED	0.153	0.231	0.201	0.159	0.142	0.127
	(0.0569)	(0.0660)	(0.0651)	(0.0512)	(0.0479)	(0.0508)
Average Standardized Scores	-0.0123	-1.111	-0.659	-0.114	0.142	0.421
	(0.448)	(0.370)	(0.404)	(0.282)	(0.270)	(0.267)
Observations	6299	204	545	2110	2501	939

Notes: Standard deviations are reported in parentheses beneath the sample means.

Tables 4 and 5 present summary statistics at the teacher level, conditional on both school and district accountability ratings, respectively. These tables only present summary statistics for my *VA* sample;[12] summary statistics for the full sample are qualitatively similar, and can be seen in Tables A.1 and A.2. Columns labeled "All" represent the summary statistics for all teachers employed at a school/district with the given accountability rating. Columns labeled "Stayers" and "Leavers" condition staying or leaving their current position, respectively. On average, school and districts with lower ratings employ teachers with less educational attainment, total experience, and contracted full-time equivalency. Moreover, schools (not districts) with lower ratings tend to have more male teachers. Conditional on school accountability rating, realized salaries seem to

[12]Methods for constructing *VA* are discussed in Section 3.3.

Table 3: School Summary Statistics, District Accountability

	Total	Fails to Meet Expectations	Meets Few Expectations	Meets Expectations	Exceeds Expectations	Sig. Exceeds Expectations
School Enrollment	370.9	482.9	414.9	340.3	375.8	429.9
	(173.5)	(179.2)	(180.8)	(161.0)	(175.7)	(172.6)
Asian Students	0.0355	0.0416	0.0394	0.0379	0.0296	0.0390
	(0.0570)	(0.0708)	(0.0676)	(0.0606)	(0.0466)	(0.0474)
Black Students	0.0903	0.525	0.370	0.0453	0.0236	0.0181
	(0.200)	(0.367)	(0.352)	(0.0696)	(0.0384)	(0.0217)
Hispanic Students	0.116	0.231	0.223	0.120	0.0711	0.0601
	(0.147)	(0.281)	(0.243)	(0.130)	(0.0688)	(0.0480)
LEP	0.0668	0.0931	0.109	0.0795	0.0399	0.0246
	(0.102)	(0.143)	(0.151)	(0.108)	(0.0576)	(0.0351)
Student Repeater	0.00408	0.00850	0.00297	0.00321	0.00290	0.0169
	(0.0367)	(0.0104)	(0.00650)	(0.0267)	(0.0239)	(0.119)
FRPL	0.457	0.768	0.758	0.487	0.328	0.181
	(0.235)	(0.176)	(0.214)	(0.179)	(0.166)	(0.148)
SPED	0.153	0.218	0.198	0.148	0.143	0.119
	(0.0575)	(0.0707)	(0.0697)	(0.0518)	(0.0473)	(0.0422)
Average Standardized Scores	-0.0127	-0.795	-0.656	-0.0299	0.223	0.500
	(0.452)	(0.481)	(0.501)	(0.298)	(0.262)	(0.220)
Observations	5366	224	558	2493	1786	305

Notes: Standard deviations are reported in parentheses beneath the sample means.

be U-shaped, as teachers at the lowest and highest rated schools receive higher wages, on average. The same is not true when conditioning on district ratings, as salaries appear to noisily increase as ratings increase.[13] Although the best rated districts and schools do have the most talented teachers, the lowest rated districts and schools do not necessarily employ the teacher's with the lowest average value-added. Lastly, turnover is highest at the lowest rated schools and districts, on average.

Tables 4 and 5 also show that stayers and leavers are different on observable characteristics. Across all ratings, leavers tend to have less educational attainment, are contracted for less FTE, and are more likely to be male. Although leavers tend to earn less, conditioning on being in the two lowest ratings show that higher paid teachers are more likely to leave their current position. The same trend is true when considering value-added, as lesser-rated schools and districts are not as successful at retaining their higher-quality teachers.

On average, if a teacher were to leave their current position, their top choice is to leave public education entirely, irrespective of accountability rating. Conditional on staying in WDPI, leavers in the lowest accountability rating are, on average, more likely to acquire new employment at a school (district) with a higher (equal) accountability rating, although the probability of that occurring is quite low. On the other hand, teachers in the lowest school and district accountability ratings are the least mobile, as they have the garner the smallest probability of leaving their district of current employment.

[13] Salary schedules are determined by districts, not schools.

Table 4: Teacher Summary Statistics, School Accountability (*VA* Sample)

	Total			Fails to Meet Expectations			Meets Few Expectations			Meets Expectations			Exceeds Expectations			Sig. Exceeds Expectations		
	All	Stayers	Leavers	All	Stayers	Leavers	All	Stayers	Leavers	All	Stayers	Leavers	All	Stayers	Leavers	All	Stayers	Leavers
Leaves current school	0.1631	0.0000	1.0000	0.2745	0.0000	1.0000	0.2287	0.0000	1.0000	0.1630	0.0000	1.0000	0.1453	0.0000	1.0000	0.1325	0.0000	1.0000
	(0.3694)	(0.0000)	(0.0000)	(0.4464)	(0.0000)	(0.0000)	(0.4200)	(0.0000)	(0.0000)	(0.3694)	(0.0000)	(0.0000)	(0.3524)	(0.0000)	(0.0000)	(0.3391)	(0.0000)	(0.0000)
Leaves to better rated school	0.0186	0.0000	0.1143	0.0611	0.0000	0.2227	0.0439	0.0000	0.1921	0.0206	0.0000	0.1262	0.0111	0.0000	0.0765	0.0035	0.0000	0.0266
	(0.1353)	(0.0000)	(0.3182)	(0.2396)	(0.0000)	(0.4164)	(0.2050)	(0.0000)	(0.3942)	(0.1419)	(0.0000)	(0.3321)	(0.1049)	(0.0000)	(0.2659)	(0.0592)	(0.0000)	(0.1609)
Leaves current district	0.1134	0.0000	0.6955	0.1689	0.0000	0.6152	0.1492	0.0000	0.6524	0.1123	0.0000	0.6885	0.1046	0.0000	0.7203	0.0987	0.0000	0.7452
	(0.3171)	(0.0000)	(0.4602)	(0.3748)	(0.0000)	(0.4870)	(0.3563)	(0.0000)	(0.4764)	(0.3157)	(0.0000)	(0.4632)	(0.3061)	(0.0000)	(0.4490)	(0.2983)	(0.0000)	(0.4360)
Leaves the WDPI	0.0812	0.0000	0.4982	0.1491	0.0000	0.5430	0.1226	0.0000	0.5363	0.0766	0.0000	0.4701	0.0727	0.0000	0.5007	0.0672	0.0000	0.5072
	(0.2732)	(0.0000)	(0.5000)	(0.3562)	(0.0000)	(0.4986)	(0.3280)	(0.0000)	(0.4989)	(0.2660)	(0.0000)	(0.4992)	(0.2597)	(0.0000)	(0.5001)	(0.2504)	(0.0000)	(0.5002)
White	0.9336	0.9389	0.9066	0.6681	0.6778	0.6426	0.7872	0.7934	0.7660	0.9471	0.9500	0.9320	0.9707	0.9711	0.9687	0.9770	0.9790	0.9638
	(0.2489)	(0.2395)	(0.2910)	(0.4710)	(0.4675)	(0.4797)	(0.4094)	(0.4049)	(0.4235)	(0.2239)	(0.2180)	(0.2518)	(0.1686)	(0.1676)	(0.1742)	(0.1501)	(0.1435)	(0.1870)
Male	0.1604	0.1587	0.1692	0.2054	0.2018	0.2148	0.1681	0.1595	0.1973	0.1592	0.1593	0.1590	0.1557	0.1546	0.1625	0.1591	0.1588	0.1606
	(0.3670)	(0.3654)	(0.3749)	(0.4041)	(0.4015)	(0.4111)	(0.3740)	(0.3662)	(0.3981)	(0.3659)	(0.3660)	(0.3657)	(0.3626)	(0.3615)	(0.3690)	(0.3658)	(0.3655)	(0.3674)
Bachelor's Degree	0.4783	0.4672	0.5350	0.6820	0.6785	0.6914	0.5759	0.5595	0.6311	0.4774	0.4691	0.5203	0.4487	0.4407	0.4958	0.4338	0.4252	0.4903
	(0.4995)	(0.4989)	(0.4988)	(0.4658)	(0.4672)	(0.4624)	(0.4943)	(0.4965)	(0.4827)	(0.4995)	(0.4991)	(0.4997)	(0.4974)	(0.4965)	(0.5001)	(0.4956)	(0.4944)	(0.5002)
Master's or Higher	0.5214	0.5326	0.4640	0.3180	0.3215	0.3086	0.4241	0.4405	0.3689	0.5220	0.5307	0.4776	0.5511	0.5592	0.5038	0.5659	0.5746	0.5085
	(0.4995)	(0.4989)	(0.4987)	(0.4658)	(0.4672)	(0.4624)	(0.4943)	(0.4965)	(0.4827)	(0.4995)	(0.4991)	(0.4996)	(0.4974)	(0.4965)	(0.5001)	(0.4957)	(0.4944)	(0.5002)
Total Salary (2012 USD / 1000)	51.9104	52.1637	50.6107	52.7125	52.3813	53.5874	52.5317	52.5279	52.5447	51.0972	51.3274	49.9151	51.9590	52.3250	49.8060	53.3432	53.6574	51.2859
	(12.6291)	(12.4271)	(13.5454)	(13.7110)	(13.7504)	(13.5807)	(13.4119)	(13.3166)	(13.7344)	(12.3774)	(12.1983)	(13.1981)	(12.4487)	(12.1999)	(13.6269)	(12.7404)	(12.5668)	(13.6538)
Total Experience	13.8698	14.0049	13.1765	11.9714	11.9303	12.0801	12.7176	12.7660	12.5546	14.0069	14.1620	13.2105	14.0543	14.1462	13.5136	14.4030	14.5495	13.4432
	(8.9821)	(8.8100)	(9.7891)	(7.9838)	(7.8997)	(8.2087)	(8.3018)	(8.1646)	(8.7506)	(9.1139)	(8.9438)	(9.9048)	(9.0629)	(8.8334)	(10.2954)	(9.0188)	(8.8859)	(9.7960)
FTE	0.9703	0.9751	0.9459	0.9797	0.9828	0.9716	0.9735	0.9779	0.9586	0.9704	0.9758	0.9428	0.9695	0.9740	0.9428	0.9671	0.9722	0.9335
	(0.1133)	(0.1045)	(0.1482)	(0.0988)	(0.0935)	(0.1114)	(0.1099)	(0.1015)	(0.1333)	(0.1120)	(0.1019)	(0.1506)	(0.1146)	(0.1064)	(0.1517)	(0.1195)	(0.1102)	(0.1643)
Math-ELA Average VA	0.0126	0.0130	0.0101	0.0136	0.0148	0.0104	0.0091	0.0087	0.0104	0.0011	0.0013	0.0003	0.0168	0.0172	0.0143	0.0345	0.0353	0.0292
	(0.0397)	(0.0392)	(0.0419)	(0.0503)	(0.0504)	(0.0499)	(0.0414)	(0.0406)	(0.0439)	(0.0368)	(0.0364)	(0.0386)	(0.0369)	(0.0363)	(0.0403)	(0.0392)	(0.0388)	(0.0416)
Observations	50960	42650	8310	1865	1353	512	5121	3950	1171	17942	15017	2925	19783	16909	2874	6249	5421	828

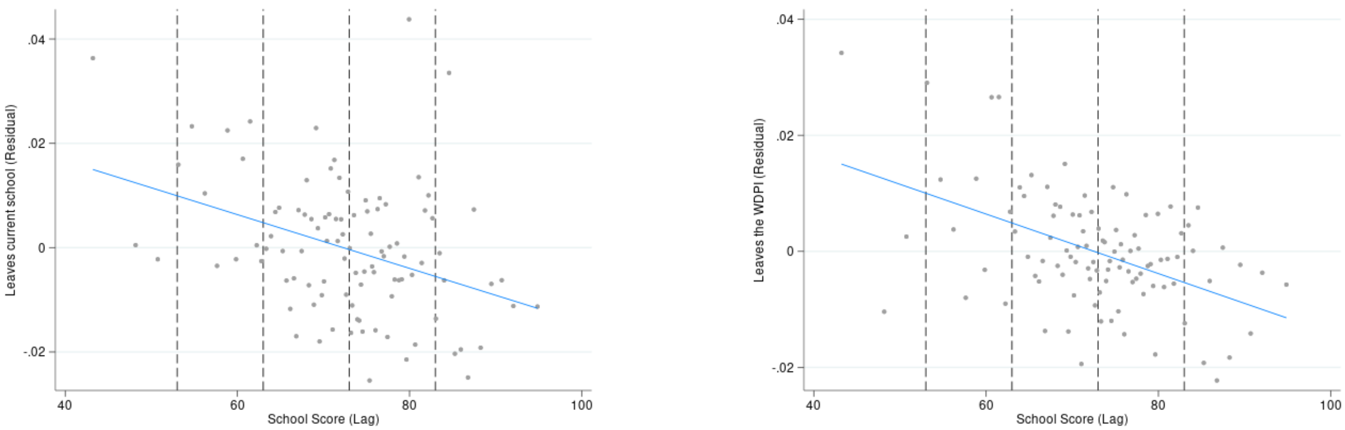
Notes: Standard deviations are reported in parentheses beneath the sample means.

Table 5: Teacher Summary Statistics, District Accountability (*VA* Sample)

	Total			Fails to Meet Expectations			Meets Few Expectations			Meets Expectations			Exceeds Expectations			Sig. Exceeds Expectations		
	All	Stayers	Leavers	All	Stayers	Leavers	All	Stayers	Leavers	All	Stayers	Leavers	All	Stayers	Leavers	All	Stayers	Leavers
Leaves current school	0.1631	0.0000	1.0000	0.2011	0.0000	1.0000	0.2894	0.0000	1.0000	0.1518	0.0000	1.0000	0.1426	0.0000	1.0000	0.1327	0.0000	1.0000
	(0.3694)	(0.0000)	(0.0000)	(0.4009)	(0.0000)	(0.0000)	(0.4535)	(0.0000)	(0.0000)	(0.3588)	(0.0000)	(0.0000)	(0.3497)	(0.0000)	(0.0000)	(0.3394)	(0.0000)	(0.0000)
Leaves current district	0.1134	0.0000	0.6955	0.1350	0.0000	0.6716	0.2285	0.0000	0.7898	0.1045	0.0000	0.6883	0.1009	0.0000	0.7072	0.1085	0.0000	0.8176
	(0.3171)	(0.0000)	(0.4602)	(0.3418)	(0.0000)	(0.4701)	(0.4200)	(0.0000)	(0.4076)	(0.3059)	(0.0000)	(0.4633)	(0.3012)	(0.0000)	(0.4552)	(0.3111)	(0.0000)	(0.3868)
Leaves to better rated district	0.0090	0.0000	0.0552	0.0090	0.0000	0.0448	0.0209	0.0000	0.0721	0.0107	0.0000	0.0706	0.0046	0.0000	0.0322	0.0013	0.0000	0.0098
	(0.0945)	(0.0000)	(0.2285)	(0.0945)	(0.0000)	(0.2070)	(0.1430)	(0.0000)	(0.2588)	(0.1029)	(0.0000)	(0.2561)	(0.0677)	(0.0000)	(0.1767)	(0.0360)	(0.0000)	(0.0985)
Leaves the WDPI	0.0812	0.0000	0.4982	0.1215	0.0000	0.6045	0.1969	0.0000	0.6803	0.0699	0.0000	0.4602	0.0685	0.0000	0.4799	0.0683	0.0000	0.5147
	(0.2732)	(0.0000)	(0.5000)	(0.3268)	(0.0000)	(0.4894)	(0.3977)	(0.0000)	(0.4666)	(0.2549)	(0.0000)	(0.4985)	(0.2525)	(0.0000)	(0.4997)	(0.2523)	(0.0000)	(0.5006)
White	0.9336	0.9389	0.9066	0.6887	0.6930	0.6716	0.8013	0.8202	0.7551	0.9588	0.9596	0.9544	0.9785	0.9791	0.9752	0.9771	0.9781	0.9707
	(0.2489)	(0.2395)	(0.2910)	(0.4631)	(0.4614)	(0.4701)	(0.3990)	(0.3841)	(0.4302)	(0.1987)	(0.1969)	(0.2087)	(0.1450)	(0.1432)	(0.1557)	(0.1497)	(0.1465)	(0.1690)
Male	0.1604	0.1587	0.1692	0.2078	0.2117	0.1922	0.1706	0.1541	0.2110	0.1505	0.1501	0.1527	0.1591	0.1578	0.1665	0.1561	0.1550	0.1629
	(0.3670)	(0.3654)	(0.3749)	(0.4058)	(0.4086)	(0.3944)	(0.3762)	(0.3611)	(0.4082)	(0.3576)	(0.3572)	(0.3597)	(0.3658)	(0.3646)	(0.3726)	(0.3630)	(0.3620)	(0.3698)
Bachelor's Degree	0.4783	0.4672	0.5350	0.6624	0.6634	0.6586	0.6148	0.5439	0.7890	0.4673	0.4651	0.4795	0.4521	0.4464	0.4863	0.4142	0.4033	0.4853
	(0.4995)	(0.4989)	(0.4988)	(0.4730)	(0.4727)	(0.4746)	(0.4867)	(0.4982)	(0.4082)	(0.4989)	(0.4988)	(0.4997)	(0.4977)	(0.4971)	(0.4999)	(0.4927)	(0.4907)	(0.5006)
Master's or Higher	0.5214	0.5326	0.4640	0.3376	0.3366	0.3414	0.3852	0.4561	0.2110	0.5322	0.5346	0.5191	0.5478	0.5535	0.5137	0.5858	0.5967	0.5147
	(0.4995)	(0.4989)	(0.4987)	(0.4730)	(0.4727)	(0.4746)	(0.4867)	(0.4982)	(0.4082)	(0.4990)	(0.4988)	(0.4997)	(0.4977)	(0.4972)	(0.4999)	(0.4927)	(0.4907)	(0.5006)
Total Salary (2012 USD / 1000)	51.9104	52.1637	50.6107	49.8634	50.4530	47.5206	54.4552	54.0977	55.3334	51.0498	51.3437	49.4075	51.2016	51.5313	49.2197	55.2638	55.8015	51.7504
	(12.6291)	(12.4271)	(13.5454)	(14.7712)	(14.9055)	(13.9944)	(12.0314)	(11.6485)	(12.8870)	(12.1970)	(12.0210)	(13.0177)	(12.3107)	(12.1213)	(13.2237)	(12.9409)	(12.7553)	(13.6003)
Total Experience	13.8698	14.0049	13.1765	12.7590	13.1223	11.3151	14.0568	14.0712	14.0214	13.9147	14.0611	13.0970	13.9102	14.0301	13.1900	14.9032	15.0725	13.7964
	(8.9821)	(8.8100)	(9.7891)	(7.7173)	(7.5381)	(8.2422)	(8.9739)	(8.8431)	(9.2914)	(9.1949)	(9.0317)	(10.0206)	(9.0505)	(8.8905)	(9.9302)	(9.0918)	(8.9486)	(9.9215)
FTE	0.9703	0.9751	0.9459	0.9810	0.9888	0.9503	0.9693	0.9702	0.9672	0.9695	0.9750	0.9389	0.9689	0.9732	0.9433	0.9755	0.9781	0.9580
	(0.1133)	(0.1045)	(0.1482)	(0.0959)	(0.0759)	(0.1472)	(0.1205)	(0.1219)	(0.1171)	(0.1141)	(0.1038)	(0.1562)	(0.1148)	(0.1071)	(0.1505)	(0.1047)	(0.0985)	(0.1377)
Math-ELA Average VA	0.0126	0.0130	0.0101	0.0210	0.0217	0.0184	0.0085	0.0072	0.0118	0.0054	0.0059	0.0023	0.0176	0.0181	0.0146	0.0366	0.0374	0.0316
	(0.0397)	(0.0392)	(0.0419)	(0.0455)	(0.0449)	(0.0479)	(0.0434)	(0.0418)	(0.0472)	(0.0382)	(0.0381)	(0.0386)	(0.0376)	(0.0369)	(0.0414)	(0.0391)	(0.0380)	(0.0454)
Observations	50960	42650	8310	2666	2130	536	3881	2758	1123	18767	15918	2849	13264	11372	1892	2313	2006	307

Notes: Standard deviations are reported in parentheses beneath the sample means.

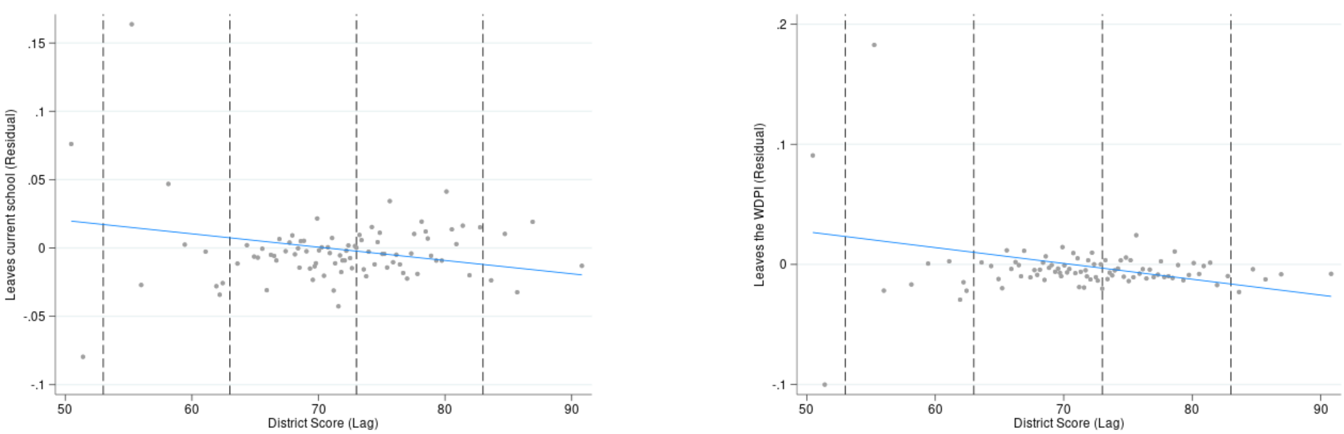
As is shown in the aforementioned tables, teachers can partially reveal their preferences for accountability scores and ratings by their employment decisions. Figures 2 and 3 provide graphical evidence to support this claim. Panel (a) of both figures show binned scatterplots of the residualized probability of a teacher leaving their current position, conditional on the realized accountability score. Panel (b) of both figures presents analogous graphical evidence, but considers the residualized probability of leaving WDPI (leaving the public school system entirely). These residuals come from regressing the respective outcomes on school, district, and academic year fixed effects. All panels indicate strong negative relationships between leaving and scores, suggesting that teachers have a distaste for poor accountabiity scores.



(a) Probability of leaving current position

(b) Probability of leaving WDPI

Figure 2: Binned scatterplot of turnover, conditional on school accountability score



(a) Probability of leaving current position

(b) Probability of leaving WDPI

Figure 3: Binned scatterplot of turnover, conditional on district accountability score

Although this paper evaluates accountability's impact on teacher labor market decisions, one potential concern is that principal and district strategy (firing, restructuring decisions) could be driving some of the effects I identify. There will certainly be turnover due to firing decisions, but for schools and districts not deemed "in need of improvement",[14] this concern can almost entirely be ruled due to the Wisconsin teacher tenure rules. Statute 119.42 prevents teachers with at least 3 years of local experience from being removed from there current position "except for cause". Moreover, Tables A.1 and A.2 show that most leavers have at least 9 years of local experience, implying that the majority of teachers left their job for reasons other than being fired.

3.3 Teacher Value-Added

Teacher value-added is a test-based measure of teacher effectiveness, conditional on other elements of achievement. In my setting, constructing a measure of teacher value-added can help reveal the determinants of the discontinuous jump in teacher turnover at each cutoff. In this section, I describe my plan for estimating teacher value-added.

3.3.1 Estimating Teacher Value-Added

Ideally, value-added would be measured by matching a class of students to their teachers, and measuring the change in standardized test scores as a result of a teacher's presence. Unfortunately, this cannot be done for the majority of my data, as WDPI began recording student-teacher linkages starting in the 2018-19 school year. For my data (including 2018-19), I can link a teacher to the students enrolled in the school and grade that they teach. To overcome this drawback, I will try to estimate teacher value-added using the methods developed by Biasi (2021). I begin with the following model of achievement:

$$
\begin{aligned}
A^*_{kt} &= \delta X_{kt} + \nu_{kt}, \\
\nu_{kt} &= \mu_{i(kt)} + \theta_{c(kt)} + \zeta_{kt},
\end{aligned}
\tag{1}
$$

where A^*_{kt} is student k's test score in year t (standardized by grade and year), X_{kt} contains observable student and school characteristics, $i(kt)$ denotes student k's teacher in t, and $c(kt)$ denotes student k's classroom. The typical procedure is to estimate the teacher-specific component of individual test score residuals, but this cannot be done in my setting.

To recover teacher value-added, Biasi (2021) suggests a grade-level approximation of teacher value-added in the spirit of Kane and Staiger (2008). Let $\tilde{i}(kt)$ refer to the set of teachers that could have taught student k at t. Also, let $g(kt)$ denote the grade of student k at t. Then $K_{igt} = \{k \mid i \in$

[14] See Section 2.1 for criteria on improvement status.

$\tilde{i}(kt),\ g(kt) = g\}$ is the set of all grade g students potentially taught by teacher i at time t. Then the average of test score residuals for all grade g students that teacher i potentially teaches at time t is:

$$\bar{A}^g_{it} = \frac{1}{N_{igt}} \sum_{k \in K_{igt}} A_{kt} \qquad (2)$$
$$A_{kt} = A^*_{kt} - \hat{\delta} X_{kt},$$

where $N_{igt} = |K_{igt}|$. Then $\hat{\mu}_i$ can be constructed in the following manner:

$$\hat{\mu}_i = \bar{v}_i \left(\frac{\sigma_u}{\text{var}(\bar{v}_i)} \right), \qquad (3)$$

where $\bar{v}_i = \sum_{g,t} w^g_{it} \bar{A}^g_{it}$ is the weighted average including each of a teacher's assigned grade level residuals. These residuals are weighted by the number of students a teacher educates in grade g at time t: $w^g_{it} = N_{igt} / \sum_{g,t} N_{igt}$. Given that these estimates are measured with noise, I include a shrinkage factor in 3, with $\sigma^2_u = \text{cov}(\bar{v}_{it}, \bar{v}_{it-1})$.

3.3.2 Identification

Given 1, identification hinges on one assumption, and one data requirement. First, I assume $E(\theta_c) = 0$, and that σ^2_c is constant. In words, I assume no peer effects in expectation. Given my data lacks classroom linkages, this assumption must be imposed to separately identify μ_i from θ_c.[15] Since I lack an estimate for θ_c, the shrinkage factor will be biased downward. However, this should not affect any qualitative results that come from my analysis.

Second, I require that there are teachers who move in my data set, which is trivially satisfied in my setting. In the absence of student-teacher linkages, the estimated residual provides a measure of the value-added for a "team" of teachers responsible for educating a given grade. If all teachers were to remain in the exact same position over my entire sample period, I would expect to see a constant team residual annually, thereby preventing me from separating an individual teacher's value-added from her team's value-added. However, given that there are teachers who move across grades and schools, and that a teacher's value-added is correlated with her team's value-added, I can utilize this variation to uniquely estimate a large portion of teachers' value added. Moreover, as the number of team combinations increases, the more teachers I can uniquely identify.

In my sample, I am able to uniquely identify the value-added for 84% of all teachers. For 12% of my sample, a teacher's value-added cannot be separately estimated from another teacher. For the

[15] The strength of this assumption can be tested with data from 2019. However, I have presently failed to perform this exercise. Future iterations of this project will evaluate this assumption in Section 3.3.3.

remaining 4%, a teacher's value-added cannot be uniquely identified from two or more teachers.

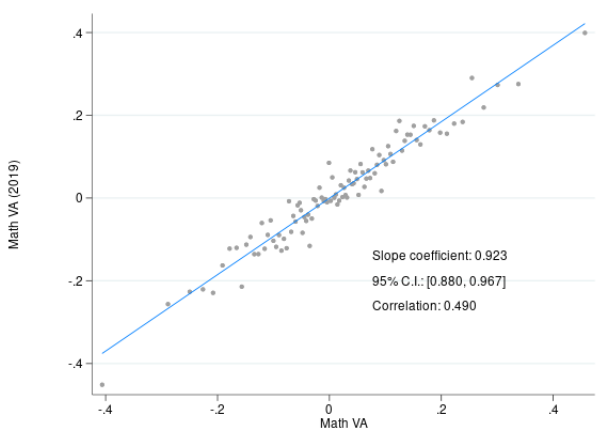
3.3.3 Forecast Bias of this Teacher Value-Added Approach

Given that I do not observe all combinations of teacher teams in my sample, there will certainly be measurement error in my value-added estimates. However, I can assess the level of measurement error by focusing on the sample of teachers in the 2018-19 school year, as the state of Wisconsin began recording student-teacher linkages during that period. For this set of teachers, I can estimate both the teacher value-added using grade level residuals (as described in Section 3.3.1; denoted as $\hat{\mu}_i^G$), and the teacher value-added using the student teacher linkages (denoted as $\hat{\mu}_i^T$). With these, I can then test if $\hat{\mu}_i^G$ is a forecast-unbiased estimate of $\hat{\mu}_i^T$ using the following model:

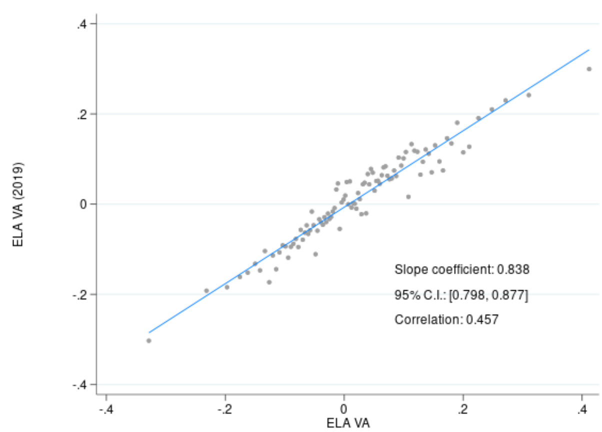
$$\hat{\mu}_i^T = \beta \hat{\mu}_i^G + \varepsilon_i. \tag{4}$$

Assuming that $\hat{\mu}_i^G$ and ε_i are uncorrelated, this test for forecast-unbiasedness has the null hypothesis that $\beta = 1$, or rather that the level of forecast bias ($f = 1 - \beta$) is zero. Figure 4 presents a binned scatterplot of the 2019 teacher value-added on the grade-level teacher value-added by subject. The mathematics value-added has a correlation coefficient of 0.49, and the confidence interval does not include 1, implying that there is a forecast bias of 0.07. The ELA value-added has a correlation coefficient of 0.45, and the confidence interval also does not include 1, implying that there is a forecast bias of 0.16.

Unfortunately, bias is present in my estimates of value-added. At this point, I proceed with my estimates, and utilize them in the design-based approach I describe in Section 4.



(a) Math *VA*



(b) ELA *VA*

Figure 4: Binned scatterplot of 2019 Teacher *VA* and Grade-level Teacher *VA*

4 Design-Based Strategy

In this section, I discuss my design-based strategy for identifying the effects of being near an accountability cutoff. For parts of my analysis, I explore effects for the full sample. In others, I condition on the value-added estimates I construct for teachers in Section 3.3.

4.1 Empirical Strategy

I adopt a regression discontinuity design (RDD) to estimate the average treatment effect of being near an accountability cutoff. For school level averages like student test scores, and the change in teacher value-added, I estimate the following model:

$$\begin{aligned} Y_{jt} &= \alpha_0 + \alpha_1 D_{jt} + \alpha_2 S_{jt} + \alpha_3 S_{jt} \cdot D_{jt} + \lambda_j + \lambda_{d(j)} + \lambda_t + \eta_{ijt} \\ D_{jt} &= \mathbb{1}\{S_{jt} \geq G\} \\ S_{jt} &\in [G-h, G+h] \\ G &\in \{53, 63, 73, 83\} \end{aligned} \tag{5}$$

In (5), j indexes schools, and t indexes school years. For school cutoffs, I estimate the effects at the four grade cutoffs (at 53, 63, 73, and 83) in four separate regressions, as denoted by G. For district cutoffs, I only estimate the top three cutoffs (63,73, and 83).[16] The running variable S_{jt} denotes school j's numeric score received in school year t. In all specifications, I include school (λ_j), district ($\lambda_{d(j)}$), and year (λ_t) fixed effects, which account for unobserved differences across schools and districts, and shocks across years, respectively. The treatment variable is $D_{jt} = \mathbb{1}\{S_{jt} \geq G\}$, which is one if a school receives an accountability score greater than cutoff G. The dependent variable is Y_{jt}, which includes student standardized test score performance, and changes in teacher value-added. The bandwidth is denoted by h. My parameter of interest is α_1, which can be interpreted as the average treatment effect at one of the four cutoffs.

When evaluating how accountability scores affect teachers' employment decisions, I estimate

[16] There are very few districts near the lowest cutoff, and estimating an RDD at this cutoff would require unreasonably large bandwidths.

the following model:

$$\begin{aligned}Y_{ijt} &= \alpha_0 + \alpha_1 D_{jt} + \alpha_2 S_{jt} + \alpha_3 S_{jt} \cdot D_{jt} + \lambda_j + \lambda_{d(j)} + \lambda_t + \eta_{ijt} \\ D_{jt} &= \mathbb{1}\{S_{jt} \geq G\} \\ S_{jt} &\in [G-h, G+h] \\ G &\in \{53, 63, 73, 83\}\end{aligned} \tag{6}$$

The model described in (6) adopts the same notation as (5), except that i indexes teachers. The dependent variable in this model is $Y_{ijt} \in \{0,1\}$, which is 1 if a teacher chooses to leave their current employment in $t+1$, and zero otherwise. Aside from estimating this specification on the full sample, I also evaluate this design on teachers with $VA > 0$, $VA < 0$, and the top 25% and bottom 25% of teachers in the value-added distribution.

In all specifications, I estimate potential discontinuities using the robust bias-correction RDD approach suggested by Calónico et al. (2014).[17] I apply a triangular weighting kernel in distance from the RDD cutoff, and use robust standard errors that are clustered at the district level. For consistency, tabular results are presented using a bandwidth of 2 accountability points on either side of the cutoff.[18] All graphical displays provide two figures for each specification: the left panel graphically displays the estimated ATE from the 2-point bandwidth, and the right panel plots how the ATE changes as I vary the bandwidth.

4.2 Density Evidence

In any RDD setting, it is assumed that the teacher, school, or district cannot manipulate the running variable. In the Wisconsin setting, it is unlikely that manipulation by teachers or schools is possible, as grade thresholds and weights are designed by the WDPI. Moreover, scores depend mainly on school and district performance on standardized tests, which are unlikely to be manipulated. That said, there is one channel I see that might encourage manipulation: the weighting schools and districts receive are conditional on them having the infrastructure to provide such data.[19] If a district were to "fail" to have the means to provide such infrastructure, this could serve as a way to manipulate their scores.

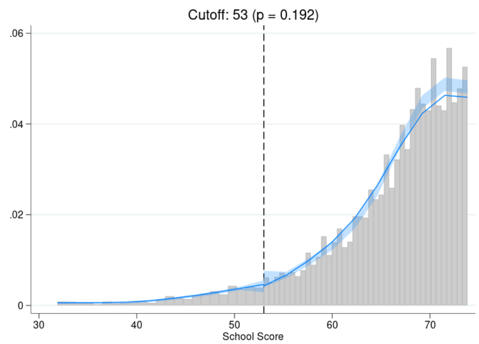
To empirically test manipulation, I perform the test proposed in Cattaneo et al. (2020). This test

[17] In my setting, "bias-correction" essentially corresponds to measuring the discontinuity given I locally estimate a quadratic line on either side of a cutoff. This approach was proposed by the aforementioned authors to account for potentially large optimal bandwidths.

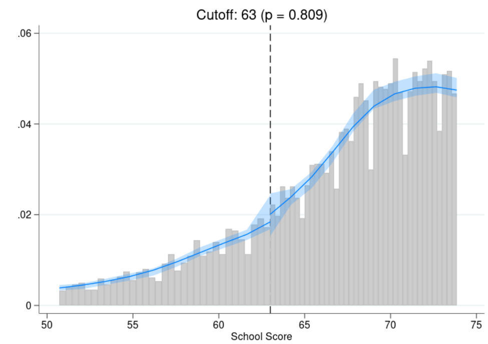
[18] When running multiple specifications, the estimated optimal bandwidths were centered around 2, reflecting this choice.

[19] To see how school/district weights vary with their availabilty of data, see https://oea-dpi.shinyapps.io/overall_weighting_calculator/

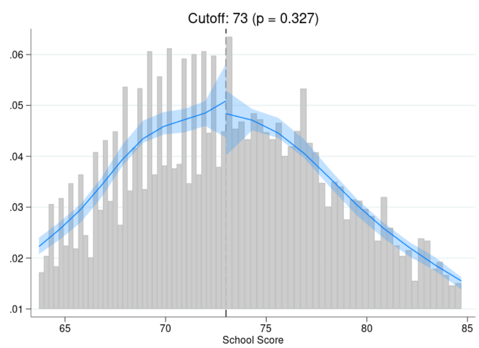
estimates potential discontinuities in the density of a running variable on either side of the cutoff using a local polynomial density estimator. If estimated discontinuities at the cutoff are statistically different than zero, inference that comes from an RDD exercise could be uninformative. Figure 5 illustrates the results of these tests for elementary schools at each school accountability cutoff. The results from these tests are statistically indistinguishable from zero, implying that I fail to reject that there is no manipulation in school accountability scores.



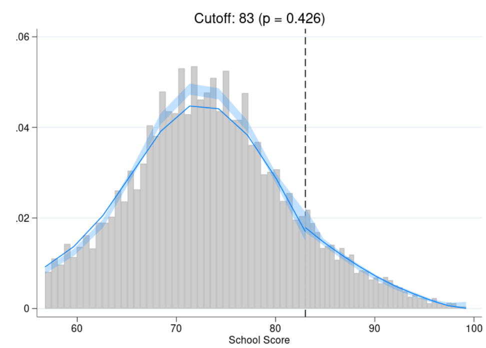
(a) Cutoff: 53



(b) Cutoff: 63



(c) Cutoff: 73



(d) Cutoff: 83

Figure 5: Density Test, Elementary Schools, by School Accountability Cutoffs

Figure A.3 displays some potential manipulation in district accountability scores at cutoff 63, with a discontinuity estimate that corresponds to a p-value of $p = 0.094$. The graphical display (in Figure A.3a) does not elicit a convincing potential discontinuity, and the inference that follows assumes that no manipulation is present. That said, it is possible that my data is "manipulated" given that only final scores are provided. WDPI provides preliminary scores to schools and districts, which they are then allowed to challenge if they notice errors in the data WDPI uses to calculate said scores. This correction may have altered the original distribution of scores. Nonetheless, there is no public documentation to support this conjecture.

4.3 Covariate Balance

Another potential threat to identification is discontinuities in the average values for covariates across the thresholds in my running variable (Imbens and Lemieux, 2008). These imbalances provide indirect evidence that the counterfactual outcome is not continuous. To test this, I treat a collection of covariates as outcomes in an RDD framework. Each observation is a school-year pair. To be consistent with my RDD analysis, I use the same bandwidth of 2 accountability points. By treating these covariates as potential outcomes in this design, I can test whether these covariates are affected by the treatment. If discontinuous jumps at the thresholds are present, this provides evidence that my results can be confounded.

Table 6: Covariate Balance, Elementary Schools, by School and District Accountability Cutoffs

	School Accountability Score				District Accountability Score		
	Cutoff: 53	Cutoff: 63	Cutoff: 73	Cutoff: 83	Cutoff: 63	Cutoff: 73	Cutoff: 83
School Enrollment	195.4581	-1.2020	-16.8771	-12.5558	38.8658	7.5859	23.5823
	(124.6881)	(62.8284)	(32.3140)	(45.9130)	(62.6635)	(33.5526)	(54.7092)
Average Attendance Rate	-0.0327	0.0147	-0.0013	-0.0018	-0.0354***	-0.0006	-0.0063
	(0.0452)	(0.0113)	(0.0062)	(0.0045)	(0.0068)	(0.0031)	(0.0050)
Male	-0.0023	-0.0055	-0.0009	0.0059	-0.0051	0.0041	0.0101
	(0.0175)	(0.0092)	(0.0055)	(0.0115)	(0.0155)	(0.0068)	(0.0108)
LEP	-0.1750	0.0131	-0.0019	0.0182	0.1820***	-0.0147	-0.0113
	(0.2896)	(0.0445)	(0.0174)	(0.0126)	(0.0275)	(0.0174)	(0.0172)
Student Repeater	0.0066	0.0006	-0.0017	0.0001	0.0011	-0.0015	-0.0072
	(0.0079)	(0.0021)	(0.0012)	(0.0014)	(0.0017)	(0.0017)	(0.0050)
FRPL	0.1460	-0.1242	-0.0347	0.0077	0.0148	-0.0371	0.0099
	(0.1676)	(0.0785)	(0.0364)	(0.0497)	(0.0492)	(0.0413)	(0.0574)
SPED	0.0539	-0.0200	0.0089	0.0125	0.0513***	0.0230*	-0.0074
	(0.0656)	(0.0183)	(0.0090)	(0.0125)	(0.0112)	(0.0093)	(0.0125)
Asian Students	0.0294	-0.0108	0.0014	0.0267	0.0746***	0.0018	0.0042
	(0.0373)	(0.0407)	(0.0142)	(0.0147)	(0.0064)	(0.0231)	(0.0151)
Hispanic Students	-0.2068	-0.0030	-0.0113	0.0185	0.1600***	-0.0327	-0.0138
	(0.3125)	(0.0633)	(0.0233)	(0.0118)	(0.0413)	(0.0198)	(0.0257)
Black Students	0.5597	-0.1464	-0.0138	-0.0422	0.0320	-0.0088	0.0025
	(0.4587)	(0.1224)	(0.0127)	(0.0470)	(0.0280)	(0.0119)	(0.0109)
Asian Teachers	0.0243	-0.0100	0.0045	-0.0016	0.0187***	0.0056	0.0005
	(0.0200)	(0.0135)	(0.0034)	(0.0041)	(0.0016)	(0.0060)	(0.0015)
Hispanic Teachers	0.0580***	0.0193	-0.0081	0.0001	0.0406***	-0.0041	-0.0047
	(0.0142)	(0.0301)	(0.0054)	(0.0028)	(0.0051)	(0.0033)	(0.0030)
Black Teachers	0.2388	-0.0577	-0.0021	-0.0166	-0.0365***	0.0016	-0.0000
	(0.1690)	(0.0498)	(0.0046)	(0.0156)	(0.0079)	(0.0018)	(0.0032)
Bachelor's Degree	0.1363	0.0300	0.0058	0.0911	-0.0746	-0.0304	0.0261
	(0.1624)	(0.0879)	(0.0368)	(0.0577)	(0.0999)	(0.0459)	(0.0667)
Master's Degree	-0.1902	-0.0190	-0.0040	-0.0902	0.0741	0.0332	-0.0243
	(0.1797)	(0.0940)	(0.0371)	(0.0594)	(0.0997)	(0.0456)	(0.0666)
Male Teachers	0.1462	-0.0224	-0.0303	0.0586*	0.0086	-0.0171	-0.0471
	(0.0794)	(0.0245)	(0.0185)	(0.0228)	(0.0150)	(0.0132)	(0.0319)
Salary (2012 USD / 1000)	10.9315*	-1.0377	-0.3727	-2.1541	3.7474	0.6422	3.6749
	(4.7303)	(2.2885)	(1.1863)	(1.5757)	(1.9504)	(1.6388)	(2.5636)
Total Experience	1.4076	0.0989	-0.1317	-1.7385*	-1.3719	0.0252	-1.2159
	(1.8252)	(0.8408)	(0.5849)	(0.8391)	(1.4872)	(0.7501)	(0.9258)
Clustering	District	District	District	District	District	District	District
Bandwidth	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000
Kernel	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular

Notes: Each cell represents an RDD estimate at the given cutoff. Standard errors are in parentheses.

Results of covariate balance tests for elementary schools at all cutoffs can be seen in Table 6.

In this table, each cell corresponds to an RDD estimate, and these estimates can be interpreted as moving just to the right of the cutoff.

When considering school accountability scores as a running variable, the share of Hispanic teachers and the average salary of teachers are estimated to be statistically different than zero at the lowest cutoff (53). However, when looking at the estimated results graphically in Figure A.4, these discontinuities appear to be a result of overfitting.

At the highest cutoff (83), there are estimated discontinuous jumps in both average total teacher experience and the share of male teachers. Panels (c) and (d) of Figure A.4 appear to show potential discontinuities. Although a discontinuity in the share of male teachers is unlikely to threathen identification, a discontinuity in the average teacher experience could, as experience has been shown to affect a teacher's effectiveness (Wiswall, 2011). Care must be taken when evaluating results at this cutoff.

Table 6 appears to show a multitude of significant discontinuities at the second lowest cutoff (63). However, graphical evidence from Figure A.5 for each of these covariates verify that these estimated discontinuities are merely due to overfitting.[20] Similarly, the share of SPED students appears to be discontinuous at cutoff 73, but the graphical display in Figure A.6 is not convincing.

Through this exercise, I find that thirteen estimates are statistically significant at the 5% level. Although I provide no joint test of significance,[21] I believe that only two of them are a legitimate discontinuities (total experience and share of male teachers, school accountability, cutoff: 83), and only one could threaten my inference when analyzing school accountability RDDs (total experience, cutoff: 83). As for district accountability, covariates are balanced at each of the cutoffs.

5 Results

For consistency, tabular results are presented using a bandwidth of 2 accountability points on either side of the cutoff. All graphical displays provide two figures for each specification: the left panel graphically displays the estimated ATE from the 2-point bandwidth, and the right panel plots how the ATE changes as I vary the bandwidth.

[20] Local linear estimates of these cutoffs also show signs of overfitting.

[21] I should have!

Table 7: Test Score RDD Estimates; Running Variable: School Accountability Score; Sample: Elementary

	Cutoff: 53				Cutoff: 63				Cutoff: 73				Cutoff: 83			
	Δ Mean ELA	Δ Mean Math	Δ P25 ELA	Δ P25 Math	Δ Mean ELA	Δ Mean Math	Δ P25 ELA	Δ P25 Math	Δ Mean ELA	Δ Mean Math	Δ P25 ELA	Δ P25 Math	Δ Mean ELA	Δ Mean Math	Δ P25 ELA	Δ P25 Math
School Score (Lag)	-0.0544	0.0115	-0.1292*	0.0657	-0.0184	-0.1331***	-0.0092	-0.1459***	0.0219	0.0312	0.0043	0.0472	0.0167	0.0340	0.0655	0.0488
	(0.0365)	(0.1111)	(0.0657)	(0.1441)	(0.0452)	(0.0369)	(0.0587)	(0.0329)	(0.0294)	(0.0244)	(0.0427)	(0.0291)	(0.0384)	(0.0364)	(0.0497)	(0.0399)
Observations	146	380	146	380	462	1468	462	1468	1373	3662	1373	3662	1155	2071	1155	2071
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	District	District	District	District	District	District	District	District	District	District	District	District	District	District	District	District
Bandwidth	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000
Kernel	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 8: Test Score RDD Estimates; Running Variable: District Accountability Score; Sample: Elementary

	Cutoff: 63				Cutoff: 73				Cutoff: 83			
	Δ Mean ELA	Δ Mean Math	Δ P25 ELA	Δ P25 Math	Δ Mean ELA	Δ Mean Math	Δ P25 ELA	Δ P25 Math	Δ Mean ELA	Δ Mean Math	Δ P25 ELA	Δ P25 Math
District Score (Lag)	-0.1664*	0.1915**	-0.2580*	0.2224*	0.0393	-0.0214	0.0329	-0.0287	0.1610***	-0.0288	0.1093*	-0.0185
	(0.0650)	(0.0725)	(0.1123)	(0.0934)	(0.0469)	(0.0488)	(0.0587)	(0.0481)	(0.0337)	(0.0280)	(0.0445)	(0.0300)
Observations	968	1739	968	1739	1974	3580	1974	3580	819	1153	819	1153
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	District	District	District	District	District	District	District	District	District	District	District	District
Bandwidth	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000
Kernel	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5.1 School Standarized Test Performance

Tables 7 presents estimates of how school accountability score lags affected student test score performance in the following academic year. For each cutoff, I test four outcomes: the change in mean and twenty-fifth percentile English/Language Arts (ELA) performance, and the change in mean and twenty-fifth percentile (henceforth, P25) Math test performance. At the lowest cutoff, "failing" schools appear to change by increasing the P25 ELA test performance by 0.12 standard deviations. However, Figure A.7 seems to display a downward trend instead of a discontinuous jump. A similar story holds for both the change in mean and P25 Math scores at the second lowest cutoff.

Table 8 presents the same estimates, except the running variable is the district accountability score. Statistical estimates show that being slightly above the second lowest threshold (63) corresponds to declines in ELA scores. Figures A.10 and A.12 do not strengthen this evidence however, as the ELA scatter around the cutoff appears noisy. On the other hand, being just to the right of the cutoff appears to increase Math scores by .2 standard deviations. Figures A.11 and A.13 show some evidence of discontinuous jumps at this cutoff, and Panel (b) of both figures shows that my estimates are fairly stable as the bandwidth changes.

Lastly, all estimates at the highest cutoff in Table 8, albeit some statistically significant, do not provide any convincing graphical evidence.

5.2 School Changes in Value-Added

My second set of results, which try to measure discontinuous jumps in the percent change in the average school value-added, can be seen in Tables 9 and 10. The 63 cutoff with school scores as a running variable is the only statistically significant estimate. However, Figure A.16 is largely influenced by an outlier that is far from the cutoff. Exclusion of this influential point would likely result in a statiscally insignificant result.

5.3 Turnover (Unconditional)

I present econometric evidence evaluating how school scores affect all teachers in Table 11. For each cutoff, I test whether a threshold changes the rate at which a teacher leaves their current position and the probability a teacher leaves full-time teaching in public education. Point estimates for all outcomes and cutoffs are small and imprecise.

Table 12 present results for all teachers when district accountability scores are the running variable. At cutoff 63, I find a statistically significant estimate for teachers leaving their current position, but visual evidence from Figure A.17 shows little action at the cutoff, and varying the

Table 9: $\% \Delta VA$ RDD Estimates; Running Variable: School Accountability Score; Sample: Elementary

	Cutoff: 53	Cutoff: 63	Cutoff: 73	Cutoff: 83
School Score (Lag)	-0.0946	-0.0194**	-0.0297	-0.0076
	(0.1235)	(0.0068)	(0.0300)	(0.0064)
Observations	98	416	1376	437
Year FE	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes
Clustering	District	District	District	District
Bandwidth	2.0000	2.0000	2.0000	2.0000
Kernel	Triangular	Triangular	Triangular	Triangular

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 10: $\% \Delta VA$ RDD Estimates; Running Variable: District Accountability Score; Sample: Elementary

	Cutoff: 63	Cutoff: 73	Cutoff: 83
District Score (Lag)	-0.1955	-0.0032	0.0117
	(0.3067)	(0.0080)	(0.0110)
Observations	258	1028	190
Year FE	Yes	Yes	Yes
School FE	Yes	Yes	Yes
District FE	Yes	Yes	Yes
Clustering	District	District	District
Bandwidth	2.0000	2.0000	2.0000
Kernel	Triangular	Triangular	Triangular

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

bandwidth shows that my point estimate is fairly unstable. The only somewhat plausible result comes from the last column of Table 12, which evaluates discontinuous jumps in leaving full-time teaching in public education at the highest cutoff. Econometrically, I find a 4 percentage point (pp) decline in leaving when just eclipsing the threshold. Visually, Figure A.20 shows some potential of a discontinuous jump, and a stable point estimate is maintained if I vary the bandwidth.

5.4 Turnover (Conditional on Value Added)

I next present estimates that mimic the strategy in Section 5.3, except these results condition on teachers that have either positive or negative estimates of value-added.[22] To begin, consider Table 13, which treats school scores as the running variable, and conditions on having a negative estimate of value-added. At the second lowest cutoff, I find strong positive effects on teachers leaving their current position and leaving full-time public education, approximately 13 pp and 7 pp, respectively. Figure A.21 and A.22 verify these estimates, as they show credible discontinuous jumps. All other estimates are measured imprecisely.

[22]Estimates that condition on being in the top/bottom 25 percentile are provided in the Appendix A.2.

Table 11: Turnover RDD Estimates; Running Variable: School Accountability Score; Sample: Elementary

	Cutoff: 53		Cutoff: 63		Cutoff: 73		Cutoff: 83	
	Leaves current school	Leaves the WDPI	Leaves current school	Leaves the WDPI	Leaves current school	Leaves the WDPI	Leaves current school	Leaves the WDPI
Accountability Score (Lag)	0.0074	-0.0095	0.0011	-0.0288	-0.0065	0.0125	-0.0108	-0.0011
	(0.0174)	(0.0076)	(0.0270)	(0.0211)	(0.0130)	(0.0090)	(0.0174)	(0.0135)
Observations	10694	10694	38788	38788	82615	82615	44525	44525
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	District	District	District	District	District	District	District	District
Bandwidth	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000
Kernel	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 12: Turnover RDD Estimates; Running Variable: District Accountability Score; Sample: Elementary

	Cutoff: 63			Cutoff: 73			Cutoff: 83		
	Leaves current school	Leaves current district	Leaves the WDPI	Leaves current school	Leaves current district	Leaves the WDPI	Leaves current school	Leaves current district	Leaves the WDPI
Accountability Score (Lag)	-0.0372***	0.0020	-0.0017	0.0071	-0.0016	-0.0049	-0.0300*	-0.0251*	-0.0426***
	(0.0100)	(0.0088)	(0.0082)	(0.0182)	(0.0121)	(0.0097)	(0.0117)	(0.0113)	(0.0077)
Observations	37843	37843	37843	77568	77568	77568	26461	26461	26461
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	District	District	District	District	District	District	District	District	District
Bandwidth	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000
Kernel	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 14 treats school scores as the running variable, and conditions on having a positive estimate of value-added. Only two estimates appear statistically significant, but they are not promising visually (see Figures A.36 and A.37).

Table 15 provides results for teachers with a negative value-added estimate when the running variable is the district accountability score. Of all the statistically significant estimates only the 63 cutoff provides potential, with 15 pp declines in teachers leaving their current position, and 5 pp declines in leaving full-time public education. Figures A.23 and A.24 present the results graphically, with the latter showing a potential discontinuous jump.

Table 16 considers how district accountability scores affect teachers with positive estimates of value-added. Only the last estimate appears plausibly discontinuous, as Figure A.42 shows a fairly clean downward jump for teachers leaving full-time public education. Moreover, the point estimates are fairly stable as I vary the bandwidth. All other statistically significant estimates seem to be the result of overfitting.

In the Appendix, I provide estimates that exclude the Milwaukee public schools to evaluate if this policy has any traction outside of Wisconsin's largest city. They can be seen in Tables A.7-A.10. Overall, these estimates are noisier, and similarly not convincing.

5.5 Discussion of Results

When evaluating standardized test performance, I find limited evidence that the second lowest threshold induces improvments in Math scores if a school is in a district just to the right of the cutoff. As for changes in school value-added, results are not convincing. From a speculative

Table 13: Turnover RDD Estimates; Running Variable: School Accountability Score; Sample: $VA < 0$

	Cutoff: 53		Cutoff: 63		Cutoff: 73		Cutoff: 83	
	Leaves current school	Leaves the WDPI	Leaves current school	Leaves the WDPI	Leaves current school	Leaves the WDPI	Leaves current school	Leaves the WDPI
Accountability Score (Lag)	-0.0303	0.0111	0.1282**	0.0687*	-0.0046	0.0338	-0.0299	-0.0012
	(0.0580)	(0.0328)	(0.0447)	(0.0307)	(0.0272)	(0.0199)	(0.0465)	(0.0327)
Observations	1555	1555	6620	6620	11611	11611	3502	3502
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	District	District	District	District	District	District	District	District
Bandwidth	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000
Kernel	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 14: Turnover RDD Estimates; Running Variable: School Accountability Score; Sample: $VA > 0$

	Cutoff: 53		Cutoff: 63		Cutoff: 73		Cutoff: 83	
	Leaves current school	Leaves the WDPI	Leaves current school	Leaves the WDPI	Leaves current school	Leaves the WDPI	Leaves current school	Leaves the WDPI
Accountability Score (Lag)	0.0115	0.0596***	-0.0578	-0.0536	0.0178	0.0403*	-0.0010	-0.0077
	(0.0167)	(0.0093)	(0.0334)	(0.0273)	(0.0286)	(0.0199)	(0.0268)	(0.0202)
Observations	2255	2255	6822	6822	17377	17377	12011	12011
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	District	District	District	District	District	District	District	District
Bandwidth	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000
Kernel	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

point of view, acquiring talented teachers is a slow process. Although school districts might be responding to these scores internally, their recruitment efforts might be fruitless.

Although there exists a strong inverse relationship between school accountability scores and teacher turnover (as Figure 2 displays), turnover at these cutoffs appear relatively smooth. Elementary school teachers may generally prefer higher graded schools to lower graded schools, but just missing a better accountability rating fail to encourage movement by teachers. Some estimates at various cutoffs are significant when conditioning on value-added, but these estimates are neither clear nor consistent to make a case that teachers really responds to being near a threshold.

When treating a district's accountability score as a running variable, I similarly see from Figure 3 that turnover and scores are inversely related. There is some evidence showing that districts just missing the second lowest cutoff respond by parting ways with their teachers with low value-added. Moreover, it appears that districts at the highest cutoff might be better at retaining their high quality teachers. Nonetheless, this evidence is weak at best, given that most of these cutoffs do not induce statistically significant effects.

I speculate that the lack of enforcement of negative consequences for failing schools is the predominant reason my analysis fails to identify discontinuities. Wisconsin's laws require that a school or distict receive repeatedly poor grades for at least 3 years before WDPI can consider intervention. Given this major delay in consequences, teachers potentially do not feel pressured by these interventions, and remain in their current job.

Table 15: Turnover RDD Estimates; Running Variable: District Accountability Score; Sample: $VA < 0$

	Cutoff: 63			Cutoff: 73			Cutoff: 83		
	Leaves current school	Leaves current district	Leaves the WDPI	Leaves current school	Leaves current district	Leaves the WDPI	Leaves current school	Leaves current district	Leaves the WDPI
Accountability Score (Lag)	-0.1459***	-0.0592	-0.0513*	0.0012	-0.0305	-0.0202	0.1323**	0.1417***	0.1069*
	(0.0204)	(0.0359)	(0.0200)	(0.0317)	(0.0209)	(0.0176)	(0.0440)	(0.0348)	(0.0474)
Observations	5704	5704	5704	10127	10127	10127	1925	1925	1925
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	District	District	District	District	District	District	District	District	District
Bandwidth	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000
Kernel	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 16: Turnover RDD Estimates; Running Variable: District Accountability Score; Sample: $VA > 0$

	Cutoff: 63			Cutoff: 73			Cutoff: 83		
	Leaves current school	Leaves current district	Leaves the WDPI	Leaves current school	Leaves current district	Leaves the WDPI	Leaves current school	Leaves current district	Leaves the WDPI
Accountability Score (Lag)	-0.2705***	-0.1345***	-0.0102	0.0163	-0.0070	-0.0043	-0.0658***	-0.0545***	-0.0797***
	(0.0759)	(0.0302)	(0.0444)	(0.0329)	(0.0265)	(0.0205)	(0.0143)	(0.0126)	(0.0061)
Observations	6621	6621	6621	15952	15952	15952	7036	7036	7036
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	District	District	District	District	District	District	District	District	District
Bandwidth	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000
Kernel	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

6 Conclusion

Accountabilty cutoffs appear to not affect any of the metrics I explore in any substantial way. Changes in standardized test performance and average school value-added appear smooth. And although teachers show some aversion to school and district accountability scores, RDD estimates of being near any threshold appear unaffected by being on either side of the cutoff. There may indeed be a preference for attending a better graded school, but my strategy shows that just missing the cutoff for a better accountability rating fails to encourage teachers to move in any meaningful ways. In the future, it might be useful to explore a difference-in-differences strategy to acquire more power for these estimates.

References

Barbara Biasi. The labor market for teachers under different pay schemes. *American Economic Journal: Economic Policy*, 74(3):63–102, 2021.

Barbara Biasi, Chao Fu, and John Stromme. Equilibrium in the market for public school teachers: District wage strategies and teacher comparative advantage. Technical report, National Bureau of Economic Research, 2021.

Sebastian Calónico, Matias Cattaneo, and Rocio Titiunik. Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326, 2014.

Matias D Cattaneo, Michael Jansson, and Xinwei Ma. Simple local polynomial density estimators. *Journal of the American Statistical Association*, 115(531):1449–1455, 2020.

Raj Chetty, John Friedman, and Jonah Rockoff. Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104(9):2633–2679, 2014.

Charles T. Clotfelter, Helen F. Ladd, Jacob L. Vigdor, and Roger Aliaga Diaz. Do school accountability systems make it more difficult for low-performing schools to attract and retain high-quality teachers? *Journal of Policy Analysis and Management*, 23(2):251–271, 2004.

David J Deming and David Figlio. Accountability in us education: Applying lessons from k-12 experience to higher education. *Journal of Economic Perspectives*, 30(3):33–56, 2016.

Rebecca Dizon-Ross. How does school accountability affect teachers? *The Journal of Human Resources*, 55(1):76–118, 2018.

Li Feng, David N Figlio, and Tim R Sass. School accountability and teacher mobility. working paper 47. *National Center for Analysis of Longitudinal Data in Education Research*, 2010.

Hege Marie Gjefsen and Trude Gunnes. The effects of School Accountability on Teacher Mobility and Teacher Sorting. MPRA Paper 69664, University Library of Munich, Germany, February 2016. URL https://ideas.repec.org/p/pra/mprapa/69664.html.

Guido Imbens and Thomas Lemieux. Regression discontinuity designs: A guide to practice. *The Journal of Econometrics*, 142(2):615–635, 2008.

M. Gail Jones, Brett D. Jones, Belinda Hardin, Lisa Chapman, Tracie Yarbrough, and Marcia Davis. The impact of high-stakes testing on teachers and students in north carolina. *The Phi Delta Kappa*, 81(3):199–203, 1999.

Thomas J Kane and Douglas O Staiger. Estimating teacher impacts on student achievement: An experimental evaluation. Technical report, National Bureau of Economic Research, 2008.

Karmen Kirtley. *High stakes testing in lower-performing high schools: Mathematics teachers' perceptions of burnout and retention*. University of Colorado at Denver, 2012.

WDPI. Archived school staff: Salary, position, and demographic reports, 2021a. URL https://dpi.wi.gov/cst/data-collections/staff/published-data.

WDPI. Wisehome and wisesecure information, 2021b. URL https://dpi.wi.gov/wise/wisehome-info.

WDPI. Public all staff report, 2021c. URL https://publicstaffreports.dpi.wi.gov/PubStaffReport/Public/PublicReport/AllStaffReport.

WDPI. Accountability history, 2021d. URL https://dpi.wi.gov/accountability/historical.

WDPI. Accountability report cards, 2021e. URL https://apps2.dpi.wi.gov/reportcards/home.

WDPI. Report card guide, 2022a. URL https://dpi.wi.gov/sites/default/files/imce/accountability/pdf/Report_Card_Guide_-_2018-19_Final_10_04_19.pdf.

WDPI. Report card timeline, 2022b. URL https://dpi.wi.gov/accountability/report-cards/timeline.

Matthew Wiswall. The dynamics of teacher quality. *Available at SSRN 1911309*, 2011.

A Appendix

A.1 Figures

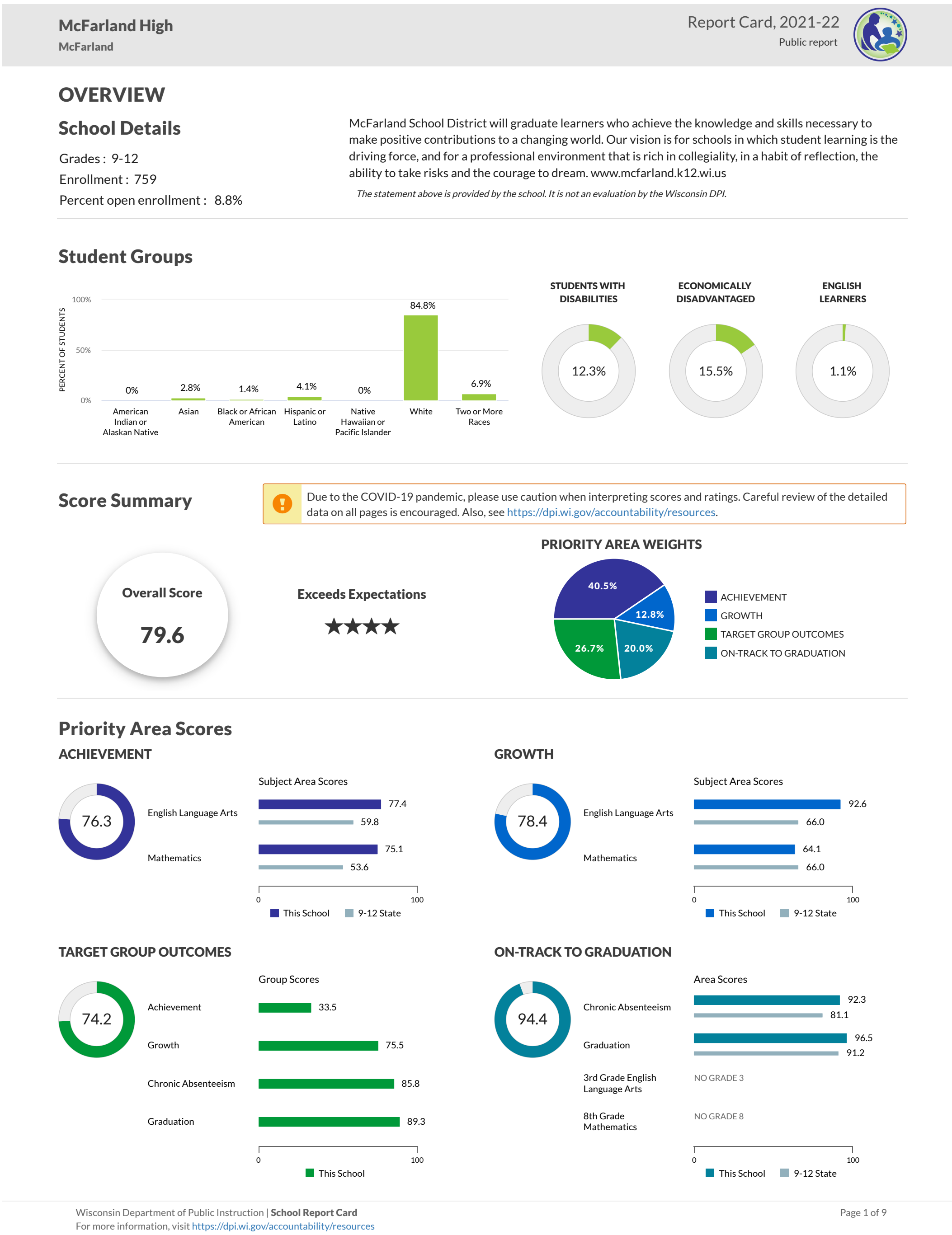
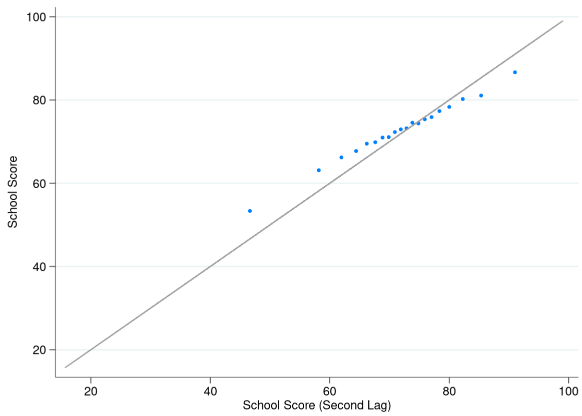
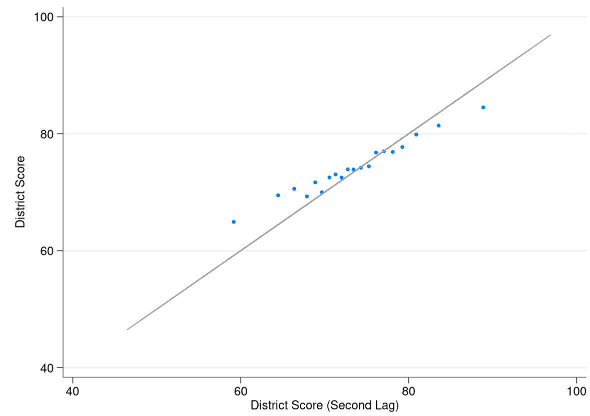


Figure A.1: Example school report card

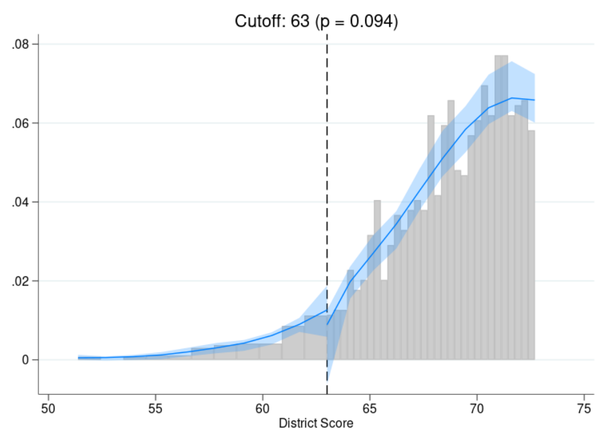


(a) School Accountability Scores

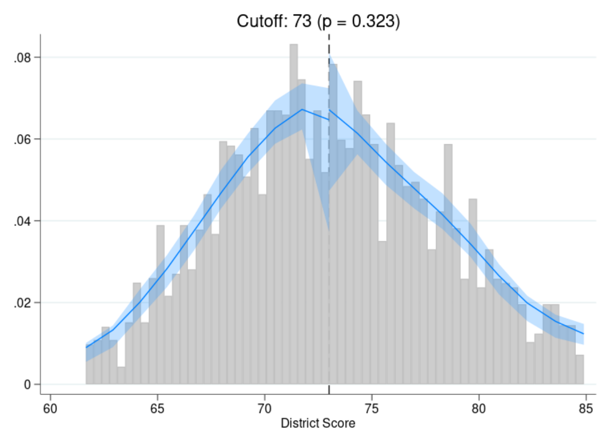


(b) District Accountability Scores

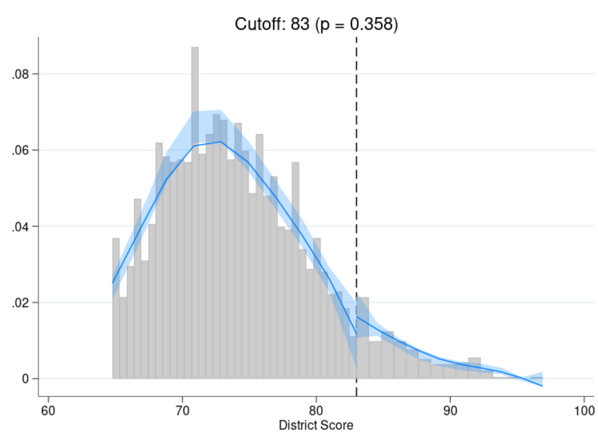
Figure A.2: Binned scatterplot of accountabilty scores, conditional on their $t-2$ score lag



(a) Cutoff: 53

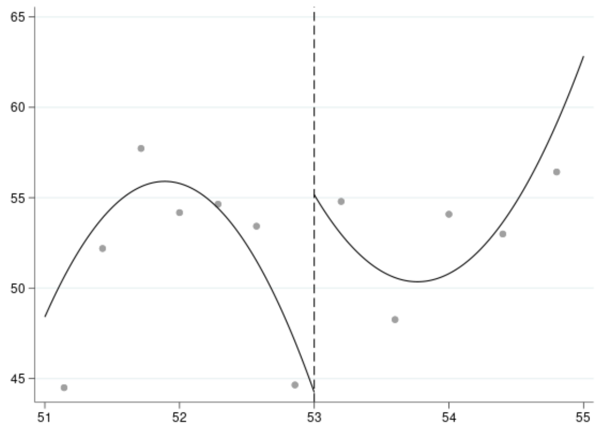


(b) Cutoff: 63

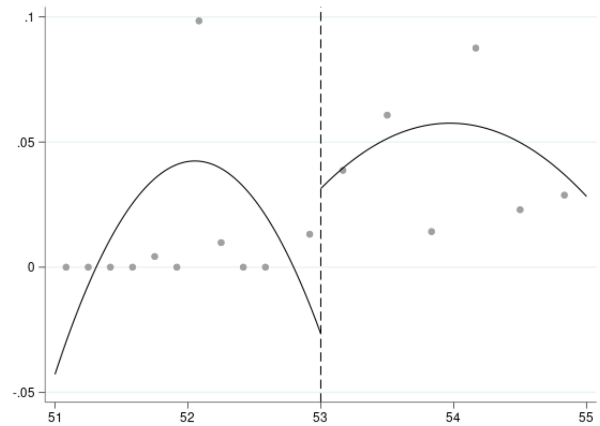


(c) Cutoff: 73

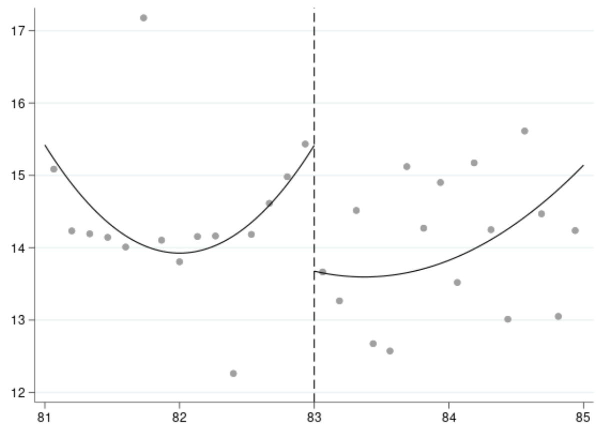
Figure A.3: Density Test, All Schools, by District Accountability Cutoffs



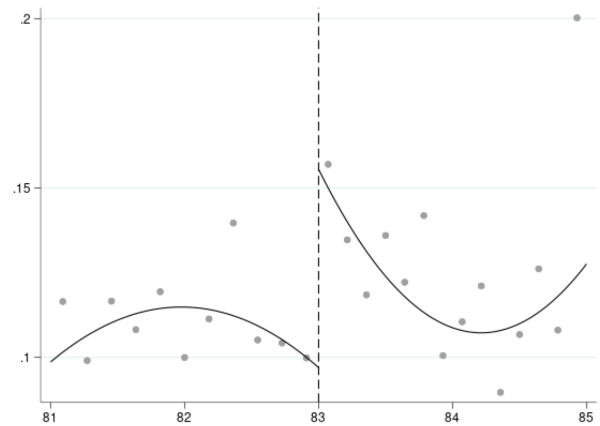
(a) Cutoff: 53; Covariate: Average Teacher Salary



(b) Cutoff: 53; Covariate: Share of Hispanic Teachers

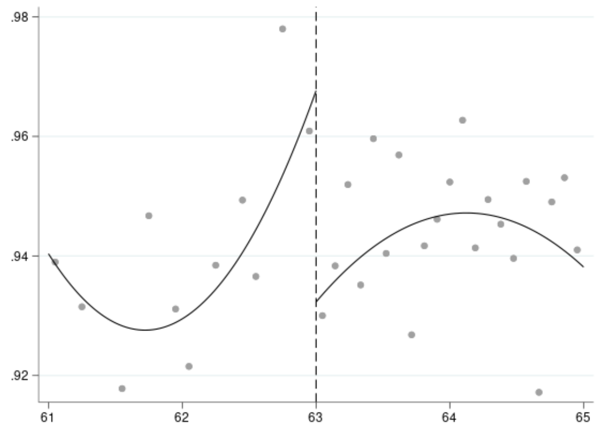


(c) Cutoff: 83; Covariate: Average Teacher Experience

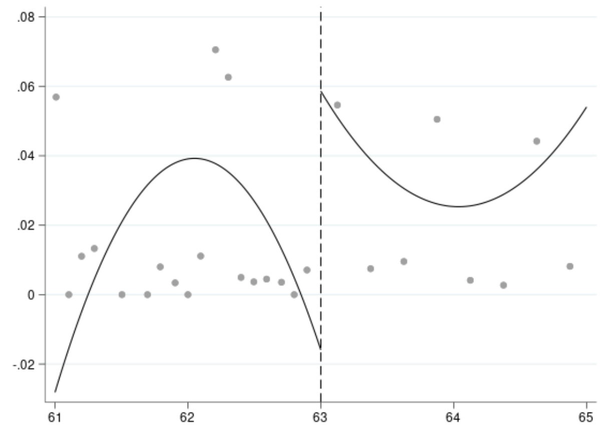


(d) Cutoff: 83; Covariate: Share of Male Teachers

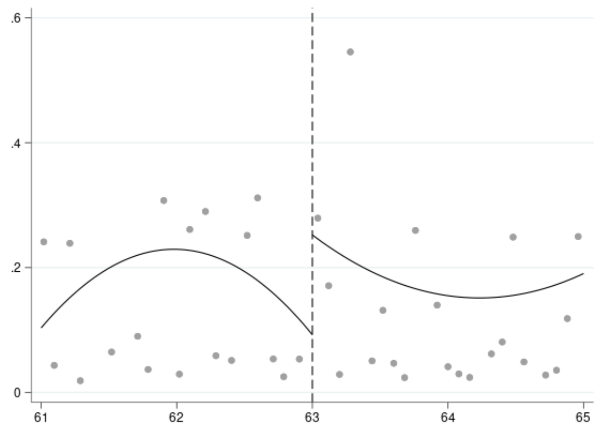
Figure A.4: Graphical View of Covariate Balance; Running Variable: School Accountability Cutoffs



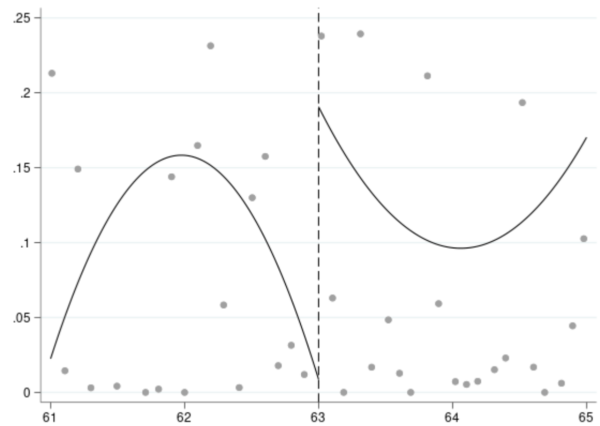
(a) Cutoff: 63; Covariate: Average Attendence Rate



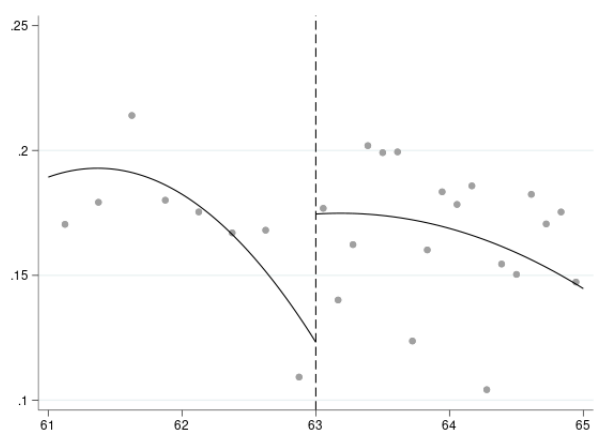
(b) Cutoff: 63; Covariate: Share of Asian Students



(c) Cutoff: 63; Covariate: Share of Hispanic Students

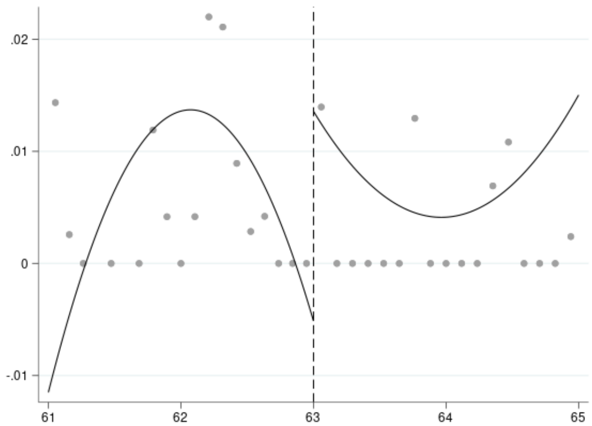


(d) Cutoff: 63; Covariate: Share of LEP Students

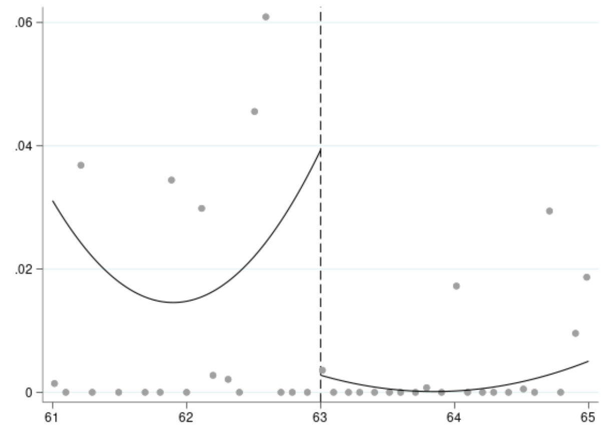


(e) Cutoff: 63; Covariate: Share of SPED Students

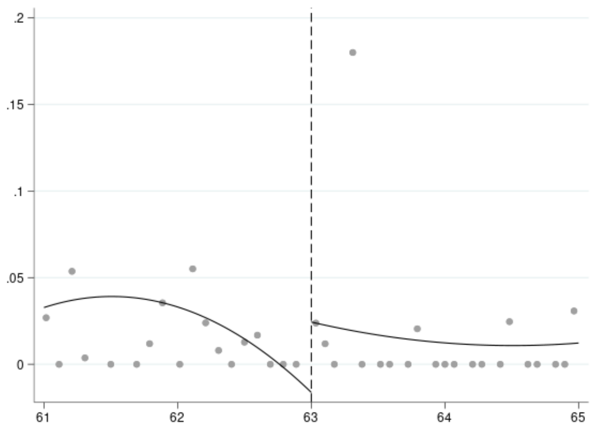
Figure A.5: Graphical View of Covariate Balance; Running Variable: District Accountability Cutoffs (Part 1)



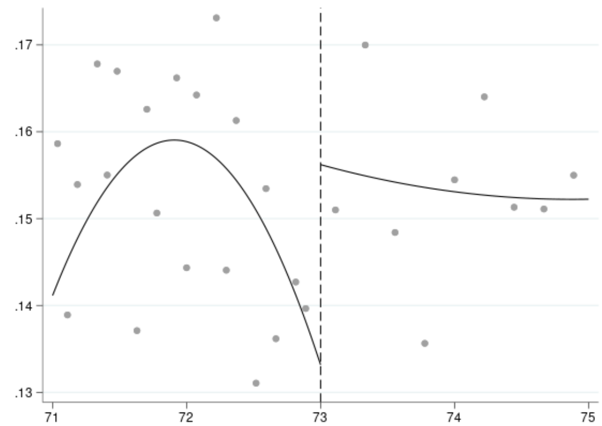
(a) Cutoff: 63; Covariate: Share of Asian Teachers



(b) Cutoff: 63; Covariate: Share of Black Teachers

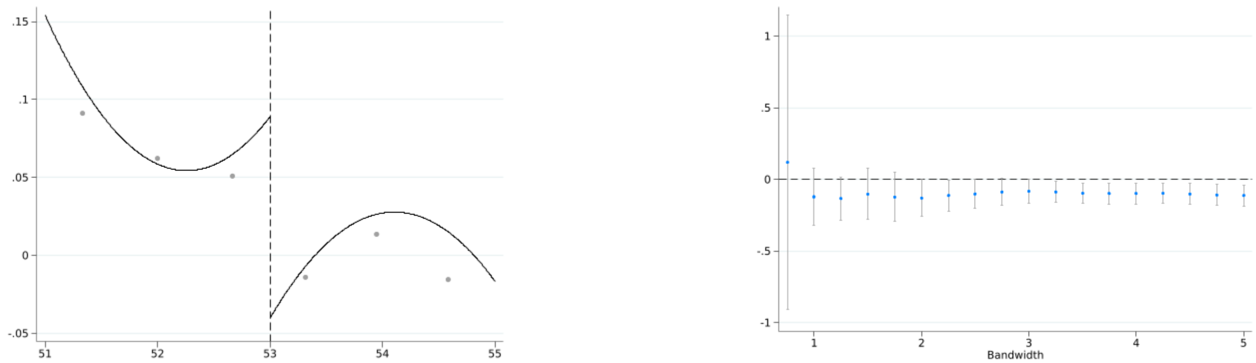


(c) Cutoff: 63; Covariate: Share of Hispanic Teachers



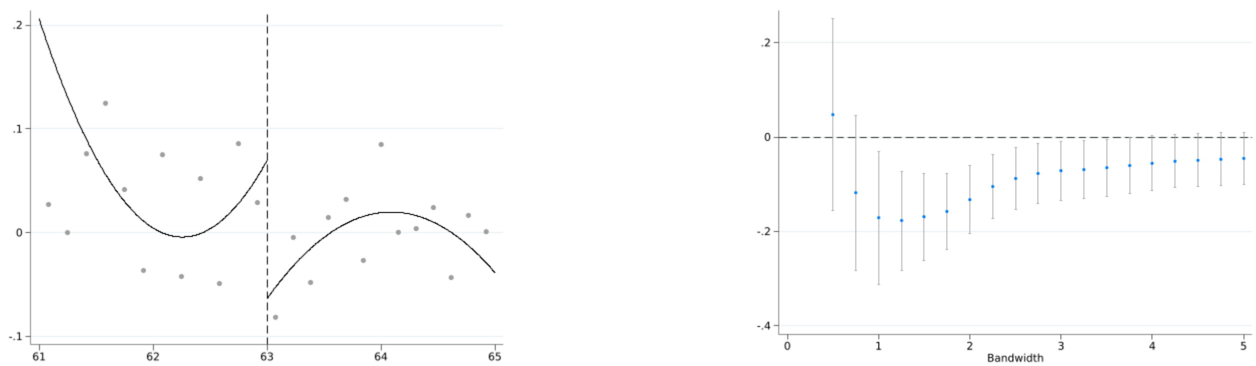
(d) Cutoff: 73; Covariate: Share of SPED Students

Figure A.6: Graphical View of Covariate Balance; Running Variable: District Accountability Cutoffs (Part 2)



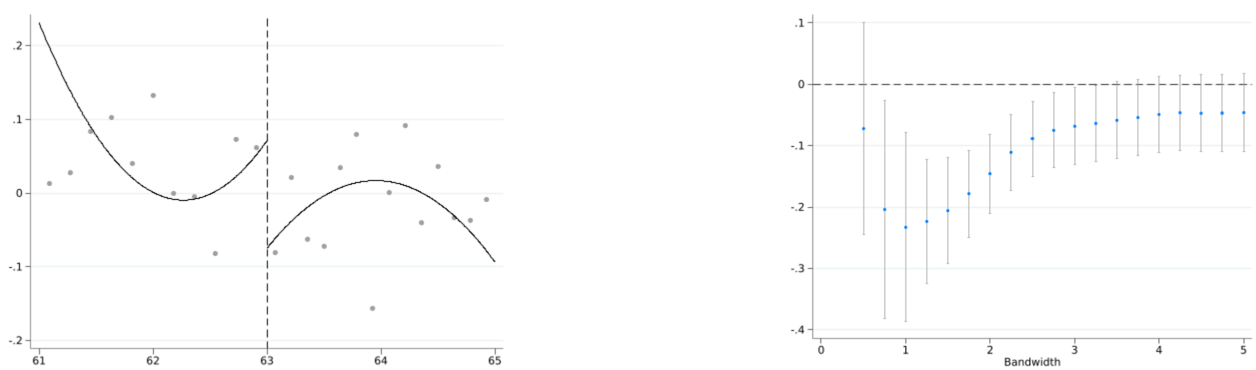
(a) RDD Plot (b) Estimate when changing bandwidth

Figure A.7: Cutoff: 53; Outcome: Δ P25 ELA; Running Variable: School Lag; Sample: Elementary



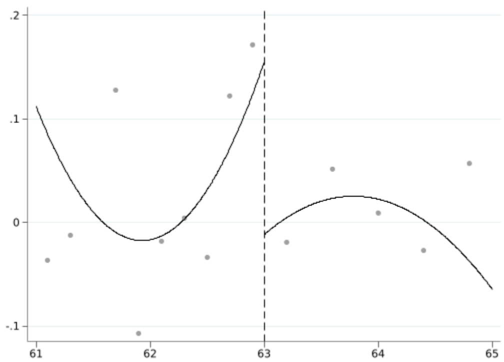
(a) RDD Plot (b) Estimate when changing bandwidth

Figure A.8: Cutoff: 63; Outcome: Δ Mean Math; Running Variable: School Lag; Sample: Elementary

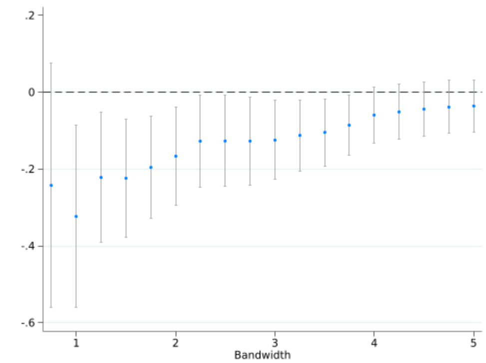


(a) RDD Plot (b) Estimate when changing bandwidth

Figure A.9: Cutoff: 63; Outcome: Δ P25 Math; Running Variable: School Lag; Sample: Elementary

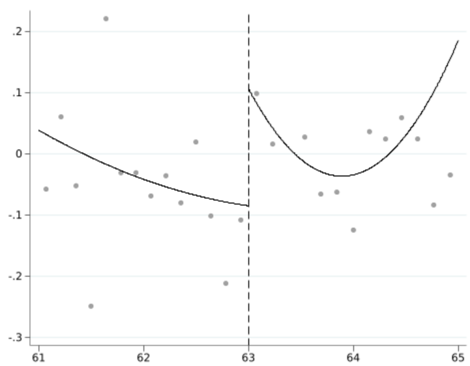


(a) RDD Plot

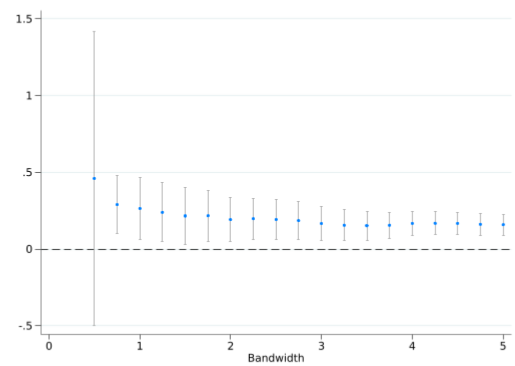


(b) Estimate when changing bandwidth

Figure A.10: Cutoff: 63; Outcome: Δ Mean ELA; Running Variable: District Lag; Sample: Elementary

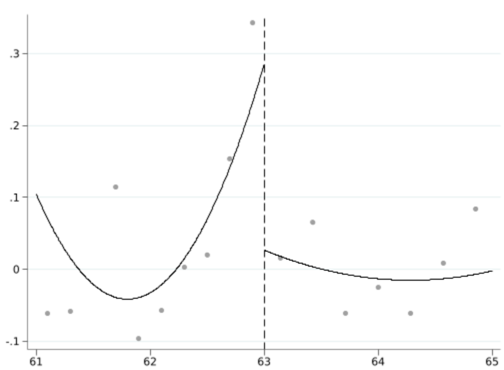


(a) RDD Plot

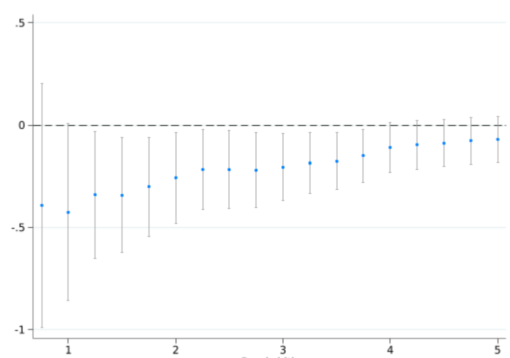


(b) Estimate when changing bandwidth

Figure A.11: Cutoff: 63; Outcome: Δ Mean Math; Running Variable: District Lag; Sample: Elementary

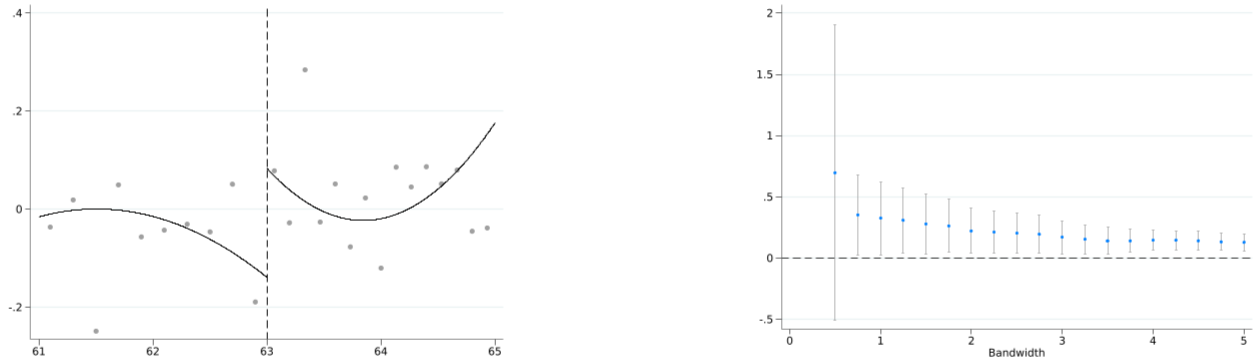


(a) RDD Plot



(b) Estimate when changing bandwidth

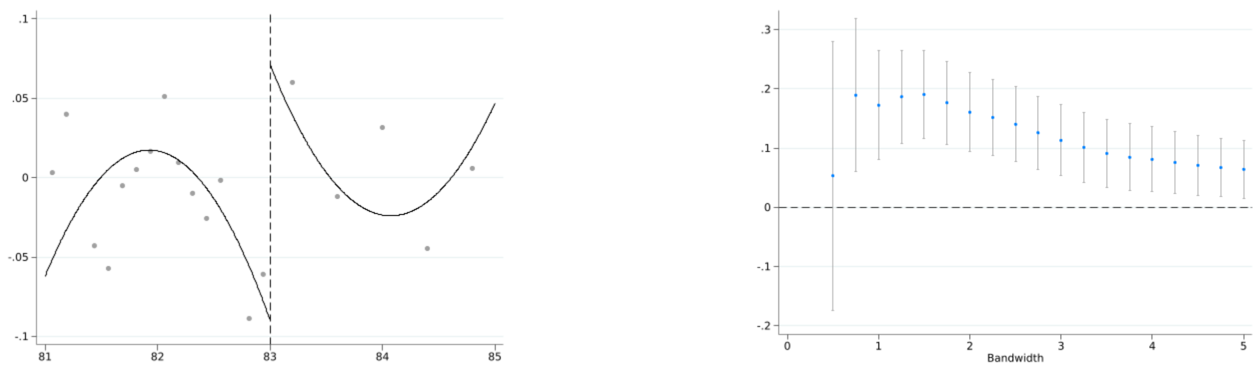
Figure A.12: Cutoff: 63; Outcome: Δ P25 ELA; Running Variable: District Lag; Sample: Elementary



(a) RDD Plot

(b) Estimate when changing bandwidth

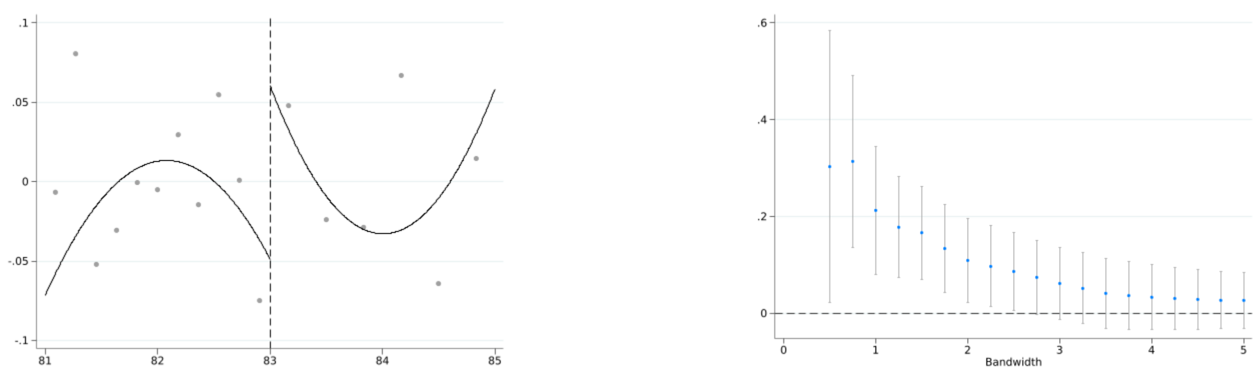
Figure A.13: Cutoff: 63; Outcome: Δ P25 Math; Running Variable: District Lag; Sample: Elementary



(a) RDD Plot

(b) Estimate when changing bandwidth

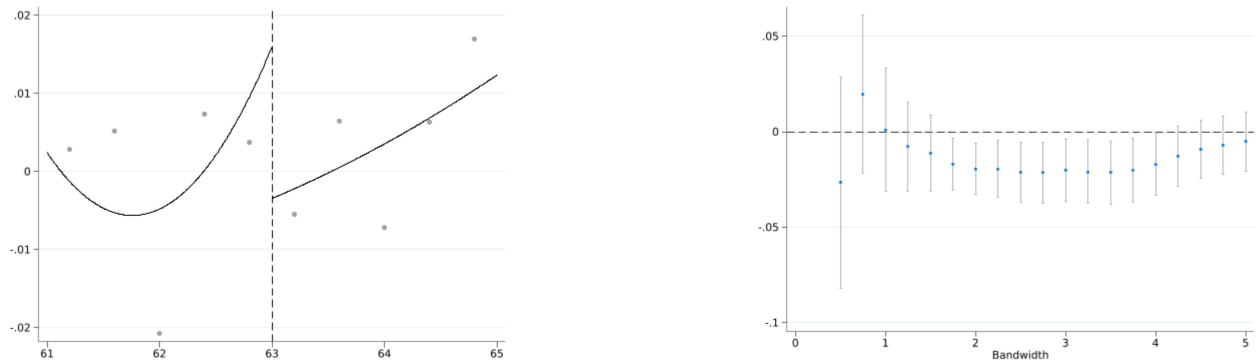
Figure A.14: Cutoff: 83; Outcome: Δ Mean ELA; Running Variable: District Lag; Sample: Elementary



(a) RDD Plot

(b) Estimate when changing bandwidth

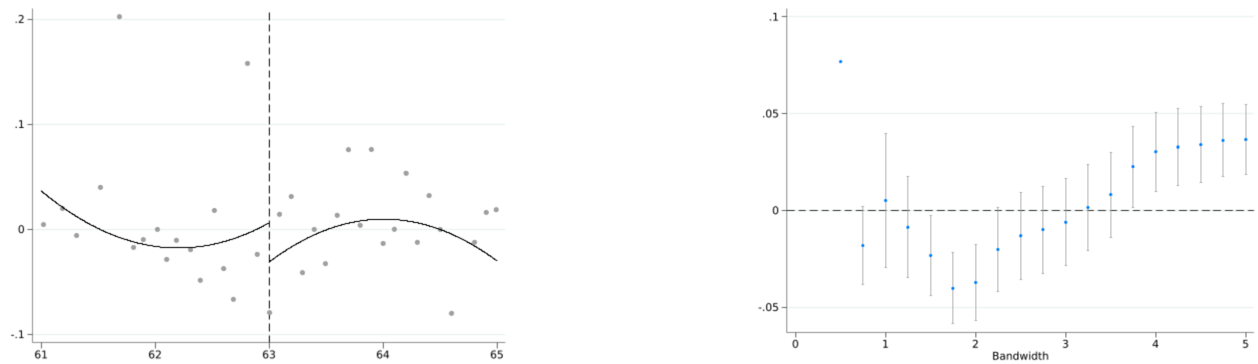
Figure A.15: Cutoff: 83; Outcome: Δ P25 ELA; Running Variable: District Lag; Sample: Elementary



(a) RDD Plot

(b) Estimate when changing bandwidth

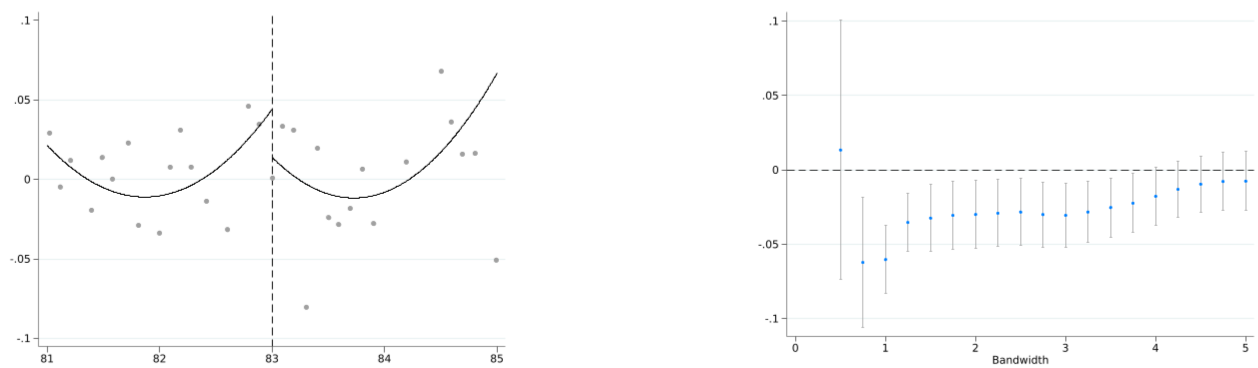
Figure A.16: Cutoff: 63; Outcome: Change in Value-Added; Running Variable: School Lag; Sample: Elementary



(a) RDD Plot

(b) Estimate when changing bandwidth

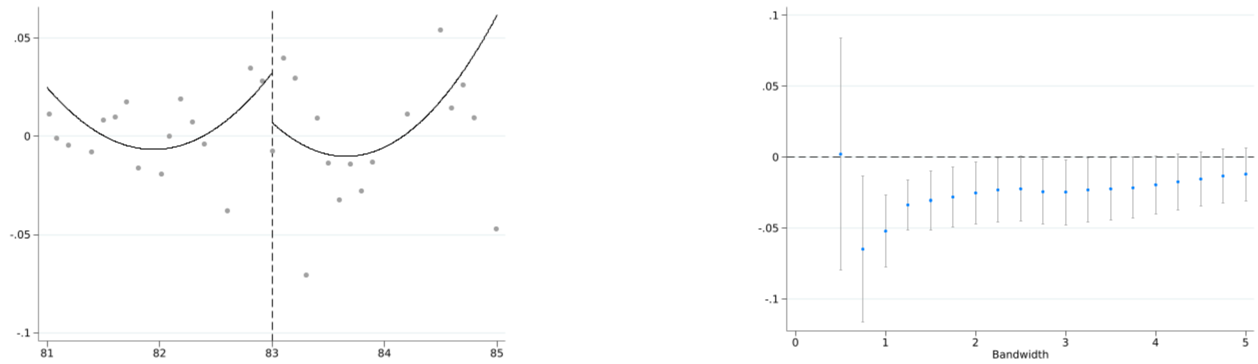
Figure A.17: Cutoff: 63; Outcome: Leaves School; Running Variable: District Lag; Sample: Elementary



(a) RDD Plot

(b) Estimate when changing bandwidth

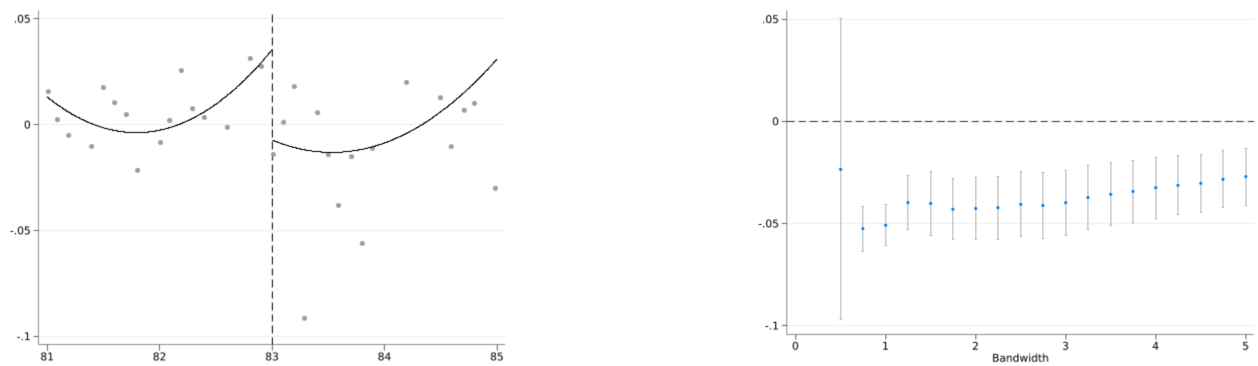
Figure A.18: Cutoff: 83; Outcome: Leaves School; Running Variable: District Lag; Sample: Elementary



(a) RDD Plot

(b) Estimate when changing bandwidth

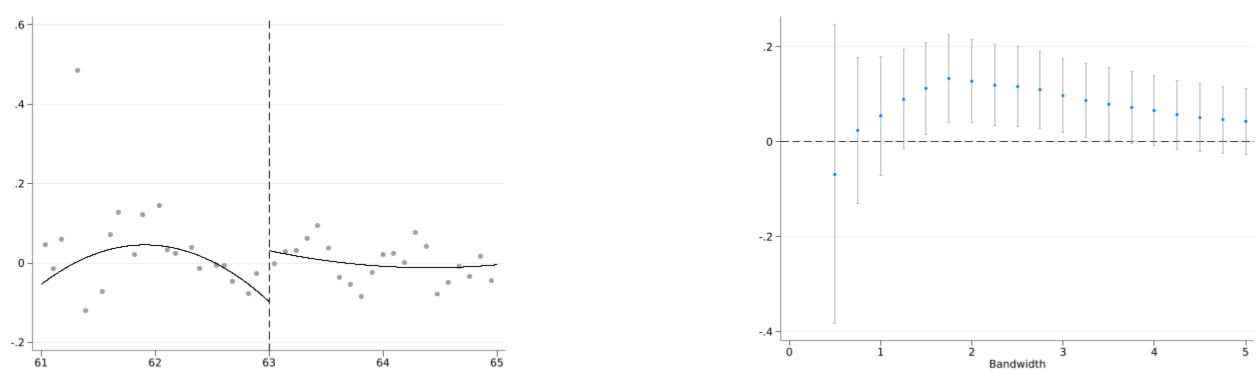
Figure A.19: Cutoff: 83; Outcome: Leaves District; Running Variable: District Lag; Sample: Elementary



(a) RDD Plot

(b) Estimate when changing bandwidth

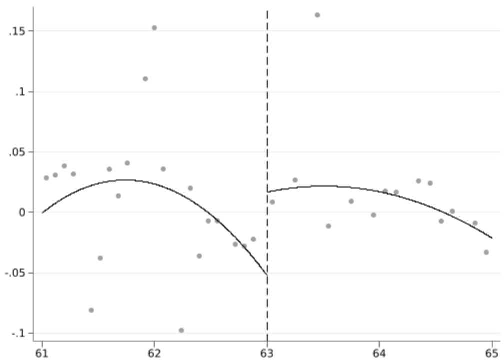
Figure A.20: Cutoff: 83; Outcome: Leaves WDPI; Running Variable: District Lag; Sample: Elementary



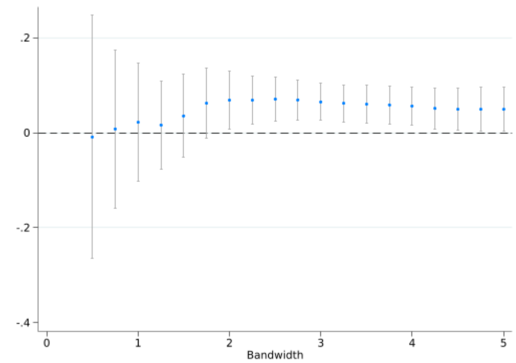
(a) RDD Plot

(b) Estimate when changing bandwidth

Figure A.21: Cutoff: 63; Outcome: Leaves School; Running Variable: School Lag; Sample: $VA < 0$

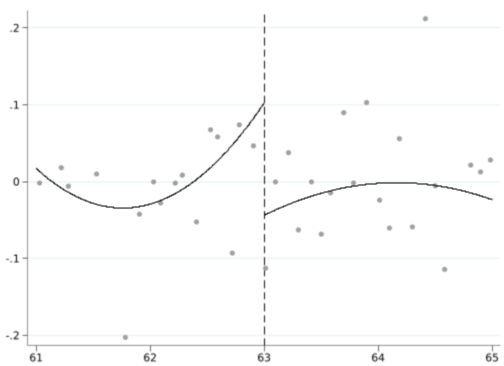


(a) RDD Plot

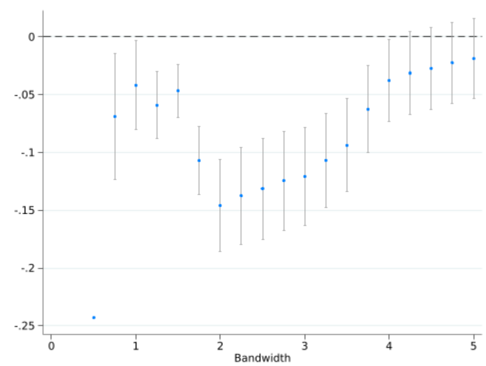


(b) Estimate when changing bandwidth

Figure A.22: Cutoff: 63; Outcome: Leaves WDPI; Running Variable: School Lag; Sample: $VA < 0$

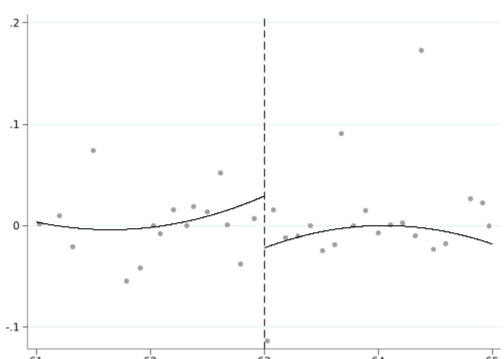


(a) RDD Plot

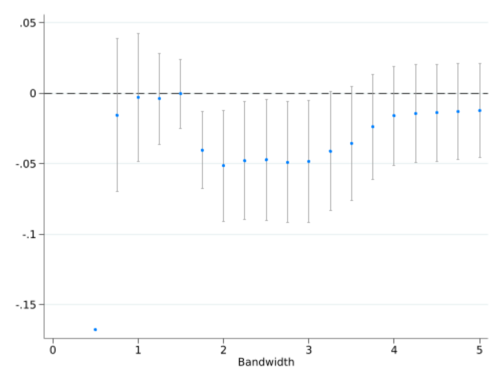


(b) Estimate when changing bandwidth

Figure A.23: Cutoff: 63; Outcome: Leaves School; Running Variable: District Lag; Sample: $VA < 0$

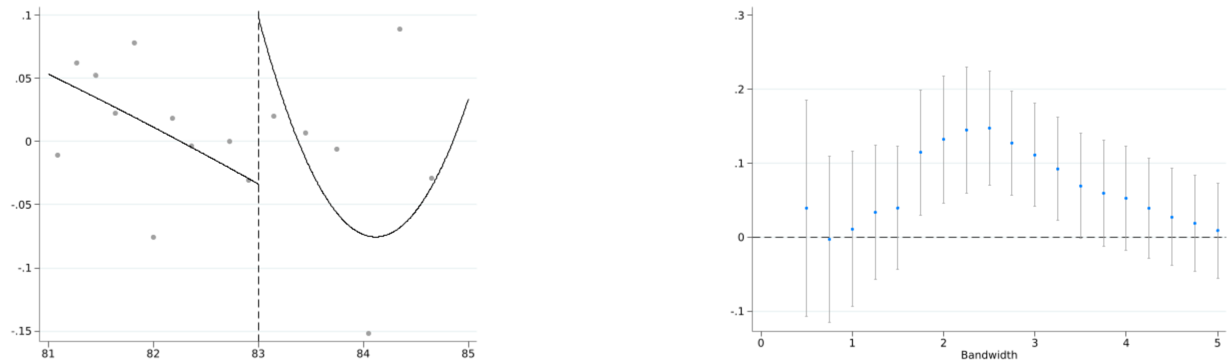


(a) RDD Plot



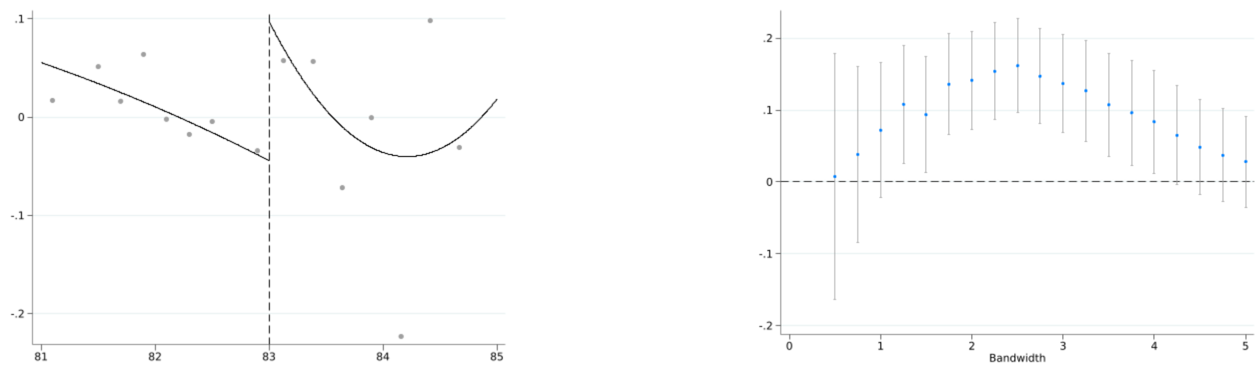
(b) Estimate when changing bandwidth

Figure A.24: Cutoff: 63; Outcome: Leaves WDPI; Running Variable: District Lag; Sample: $VA < 0$



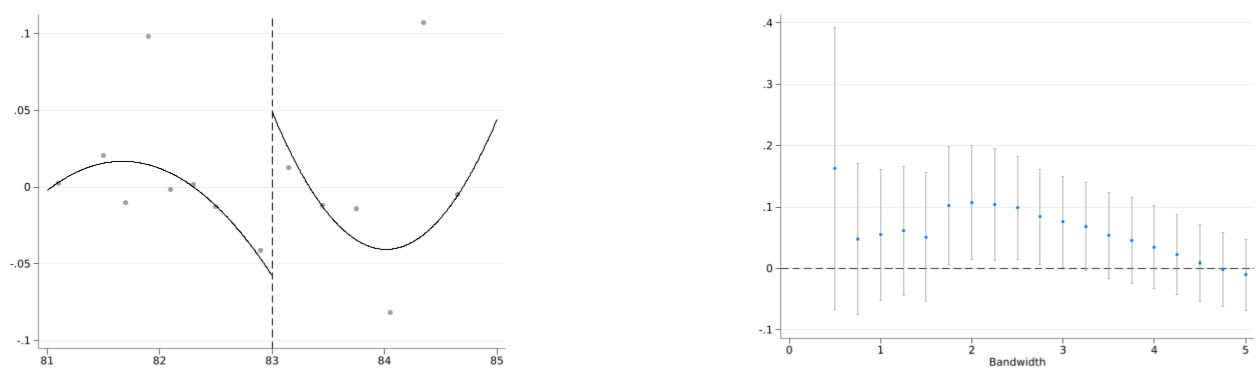
(a) RDD Plot (b) Estimate when changing bandwidth

Figure A.25: Cutoff: 83; Outcome: Leaves School; Running Variable: District Lag; Sample: $VA < 0$



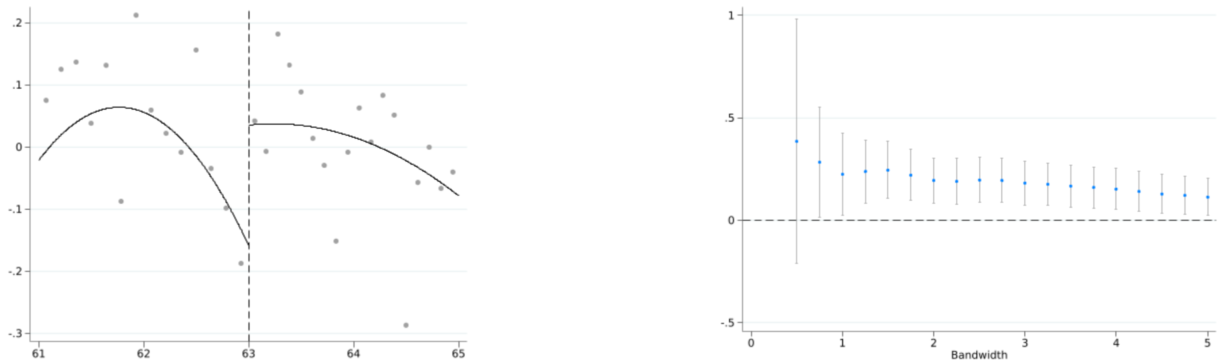
(a) RDD Plot (b) Estimate when changing bandwidth

Figure A.26: Cutoff: 83; Outcome: Leaves District; Running Variable: District Lag; Sample: $VA < 0$



(a) RDD Plot (b) Estimate when changing bandwidth

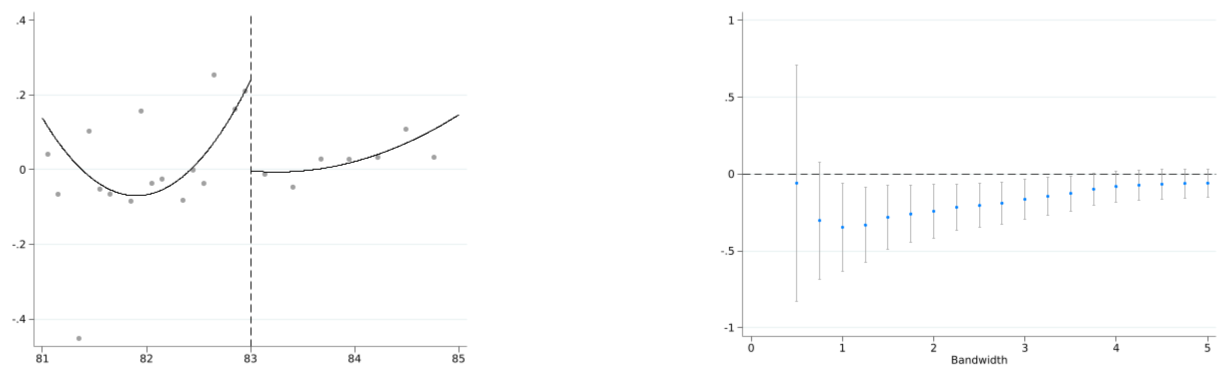
Figure A.27: Cutoff: 83; Outcome: Leaves WDPI; Running Variable: District Lag; Sample: $VA < 0$



(a) RDD Plot

(b) Estimate when changing bandwidth

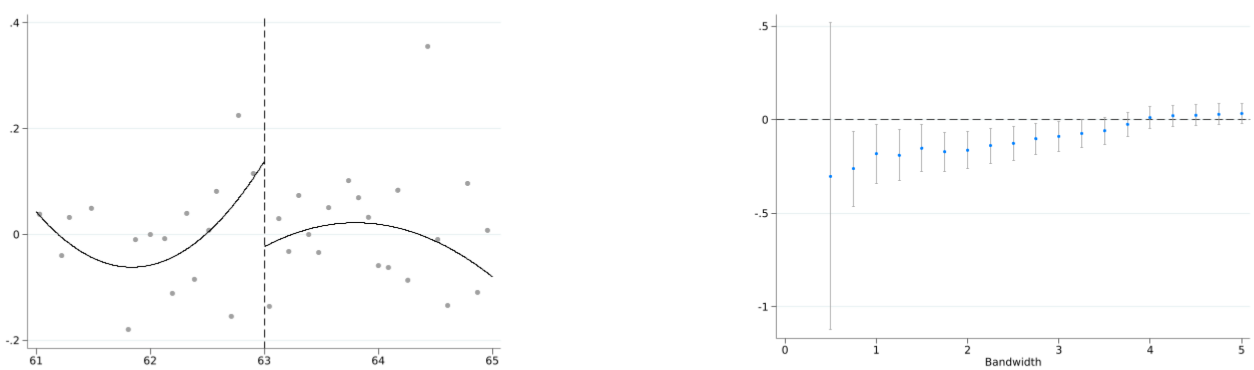
Figure A.28: Cutoff: 63; Outcome: Leaves School; Running Variable: School Lag; Sample: Bottom 25% *VA*



(a) RDD Plot

(b) Estimate when changing bandwidth

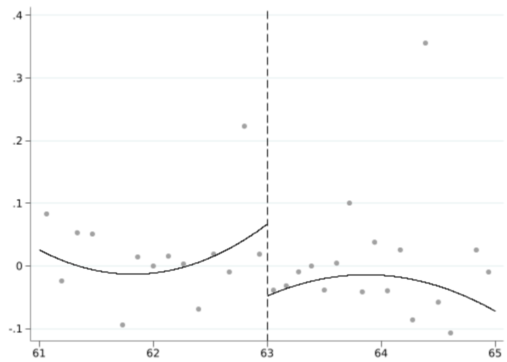
Figure A.29: Cutoff: 83; Outcome: Leaves School; Running Variable: School Lag; Sample: Bottom 25% *VA*



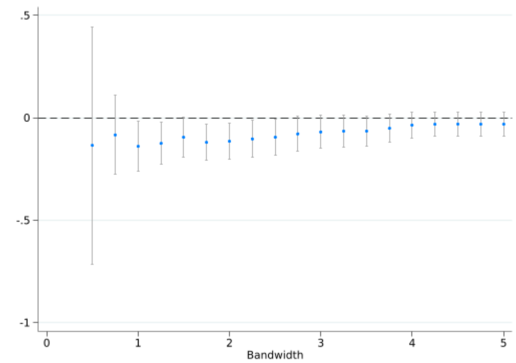
(a) RDD Plot

(b) Estimate when changing bandwidth

Figure A.30: Cutoff: 63; Outcome: Leaves School; Running Variable: District Lag; Sample: Bottom 25% *VA*

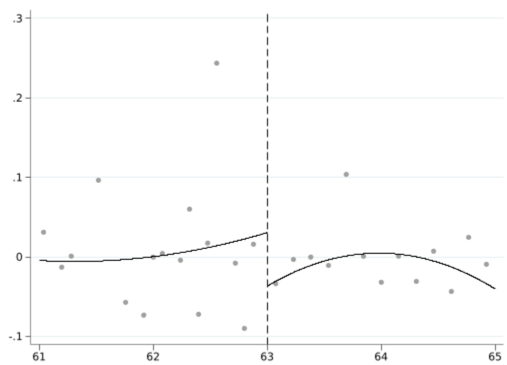


(a) RDD Plot

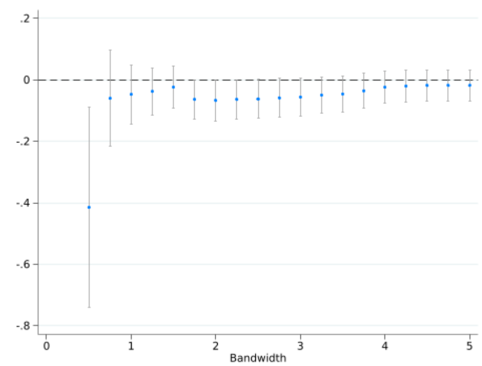


(b) Estimate when changing bandwidth

Figure A.31: Cutoff: 63; Outcome: Leaves District; Running Variable: District Lag; Sample: Bottom 25% *VA*

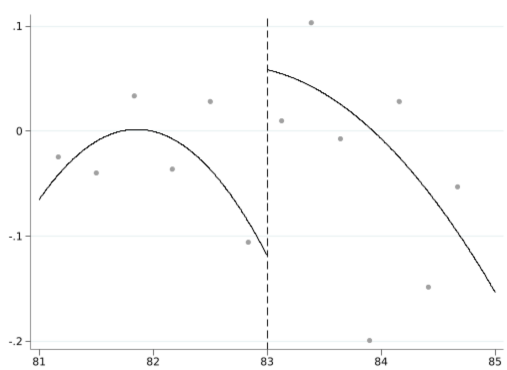


(a) RDD Plot

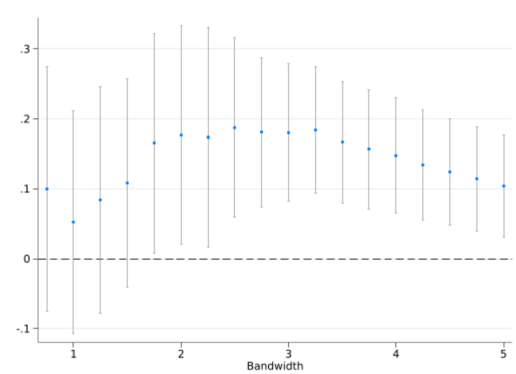


(b) Estimate when changing bandwidth

Figure A.32: Cutoff: 63; Outcome: Leaves WDPI; Running Variable: District Lag; Sample: Bottom 25% *VA*

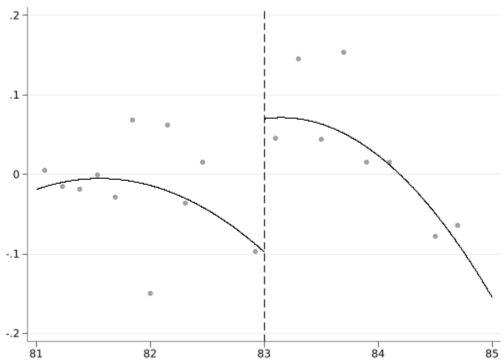


(a) RDD Plot

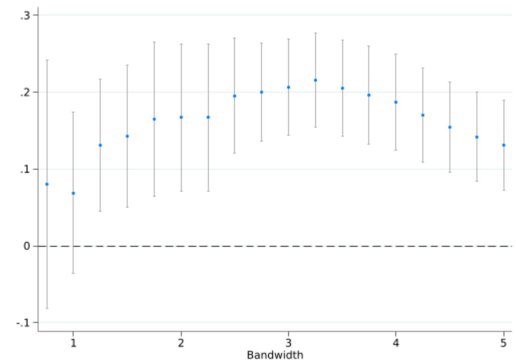


(b) Estimate when changing bandwidth

Figure A.33: Cutoff: 83; Outcome: Leaves School; Running Variable: District Lag; Sample: Bottom 25% *VA*

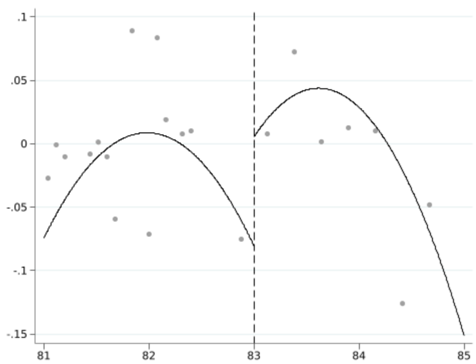


(a) RDD Plot

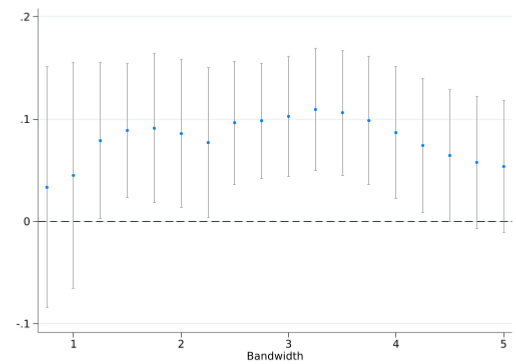


(b) Estimate when changing bandwidth

Figure A.34: Cutoff: 83; Outcome: Leaves District; Running Variable: District Lag; Sample: Bottom 25% *VA*

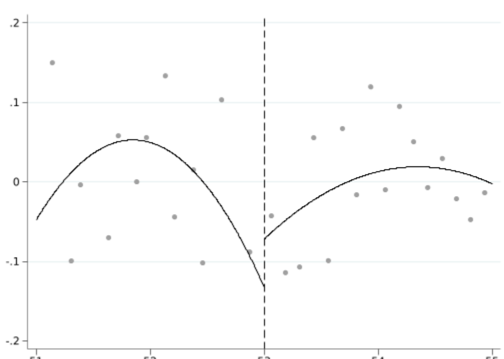


(a) RDD Plot

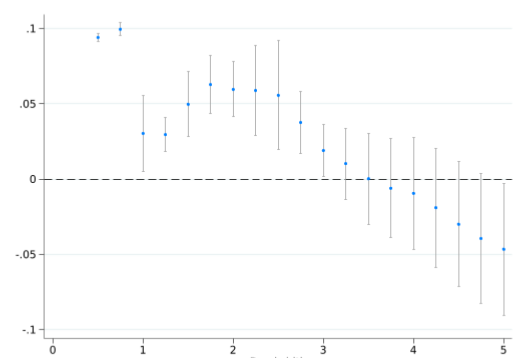


(b) Estimate when changing bandwidth

Figure A.35: Cutoff: 83; Outcome: Leaves WDPI; Running Variable: District Lag; Sample: Bottom 25% *VA*

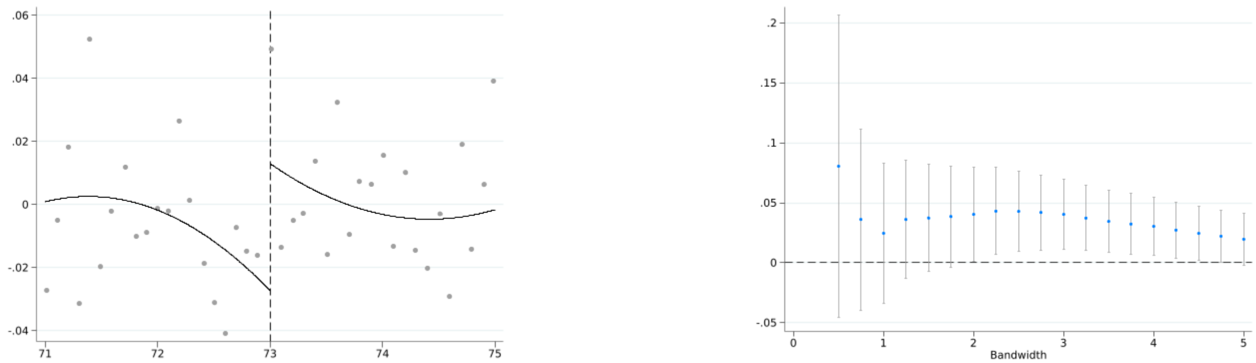


(a) RDD Plot



(b) Estimate when changing bandwidth

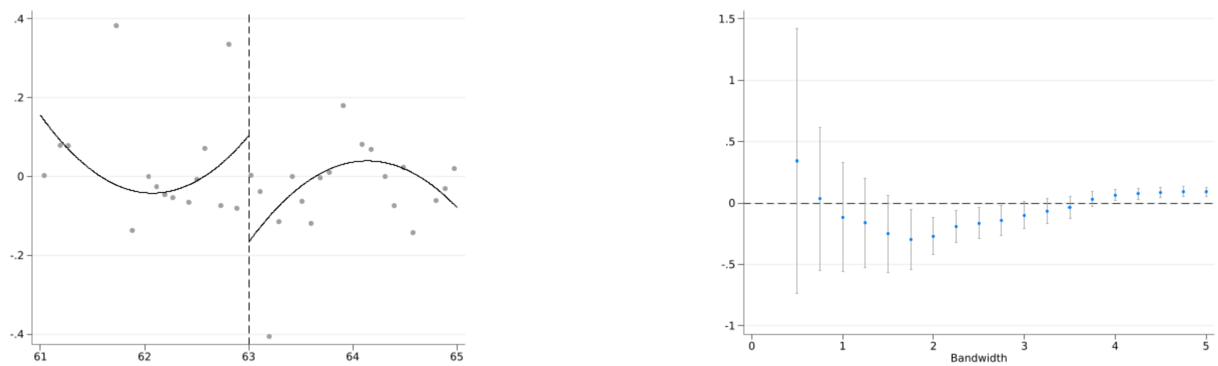
Figure A.36: Cutoff: 53; Outcome: Leaves WDPI; Running Variable: School Lag; Sample: $VA > 0$



(a) RDD Plot

(b) Estimate when changing bandwidth

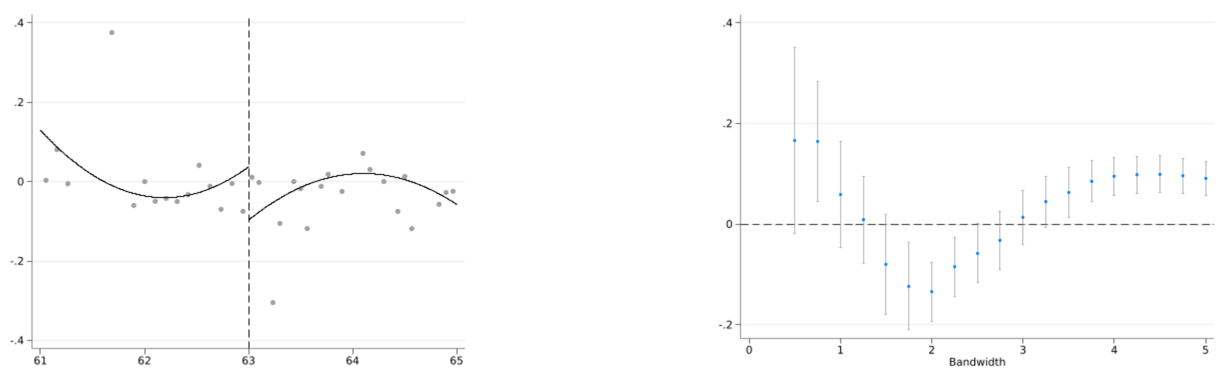
Figure A.37: Cutoff: 73; Outcome: Leaves WDPI; Running Variable: School Lag; Sample: $VA > 0$



(a) RDD Plot

(b) Estimate when changing bandwidth

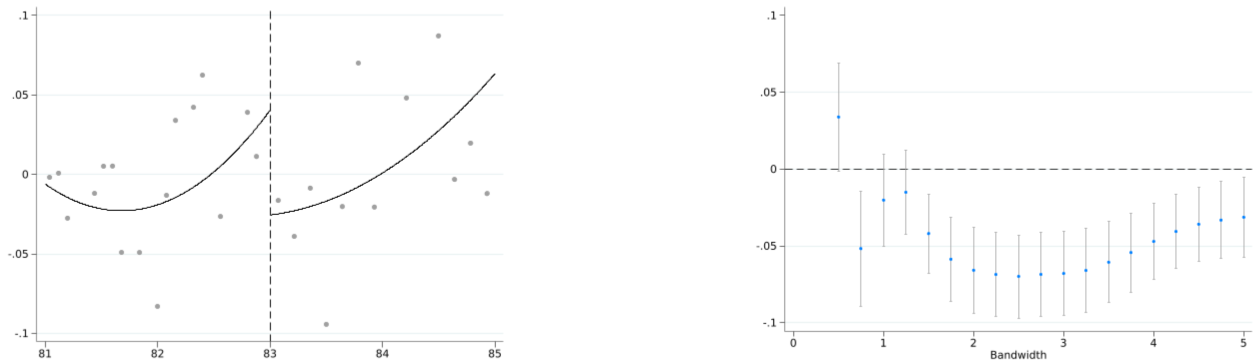
Figure A.38: Cutoff: 63; Outcome: Leaves School; Running Variable: District Lag; Sample: $VA > 0$



(a) RDD Plot

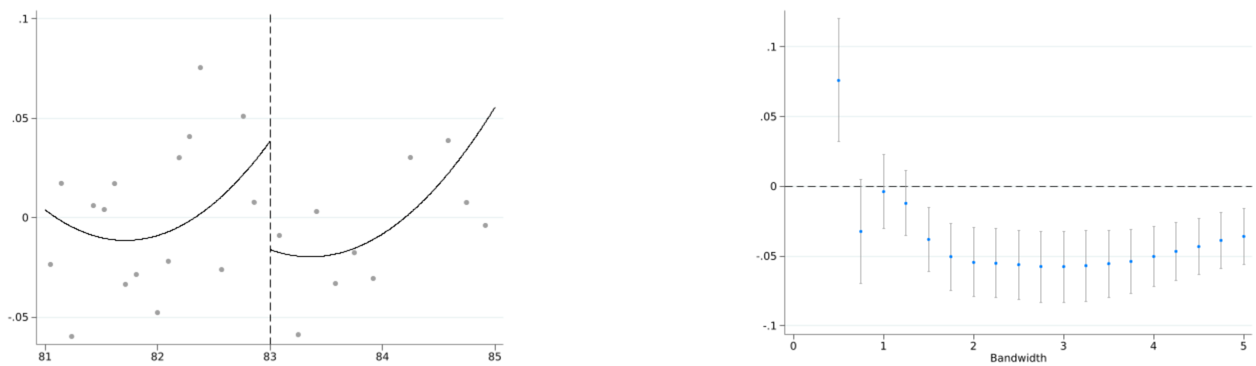
(b) Estimate when changing bandwidth

Figure A.39: Cutoff: 63; Outcome: Leaves District; Running Variable: District Lag; Sample: $VA > 0$



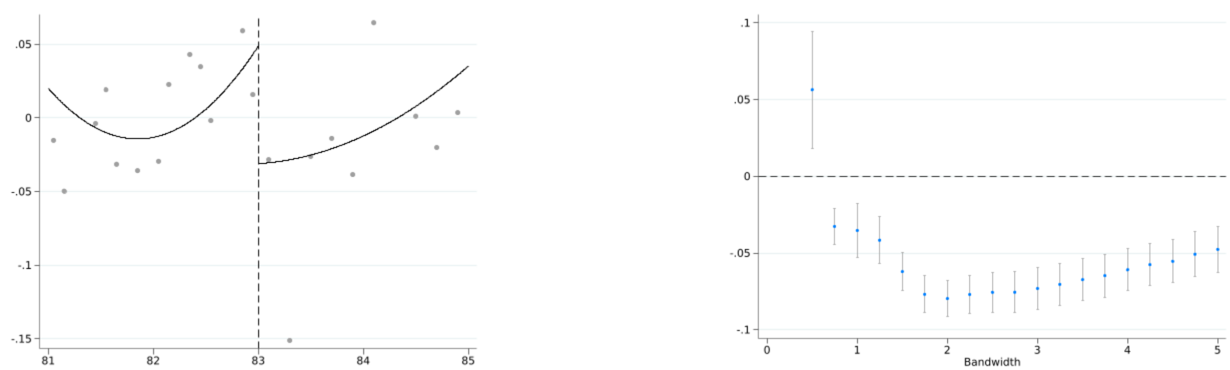
(a) RDD Plot (b) Estimate when changing bandwidth

Figure A.40: Cutoff: 83; Outcome: Leaves School; Running Variable: District Lag; Sample: $VA > 0$



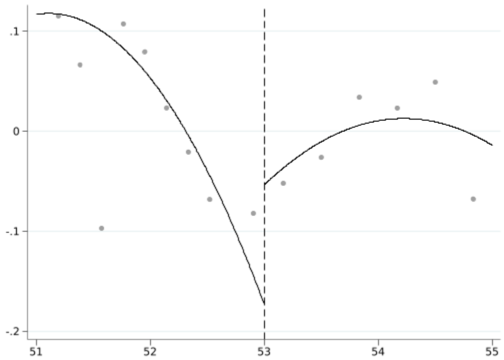
(a) RDD Plot (b) Estimate when changing bandwidth

Figure A.41: Cutoff: 83; Outcome: Leaves District; Running Variable: District Lag; Sample: $VA > 0$

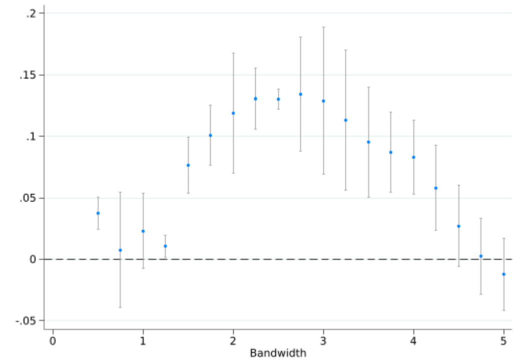


(a) RDD Plot (b) Estimate when changing bandwidth

Figure A.42: Cutoff: 83; Outcome: Leaves WDPI; Running Variable: District Lag; Sample: $VA > 0$

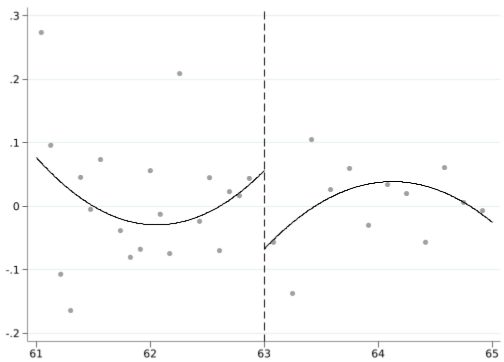


(a) RDD Plot

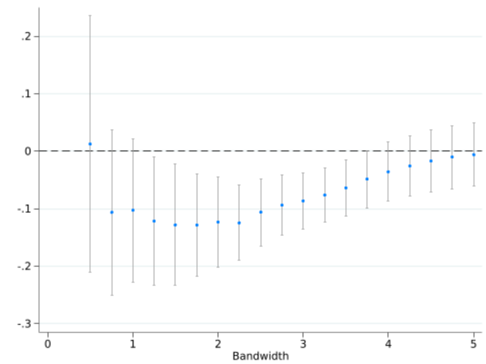


(b) Estimate when changing bandwidth

Figure A.43: Cutoff: 53; Outcome: Leaves WDPI; Running Variable: School Lag; Sample: Top 25% *VA*

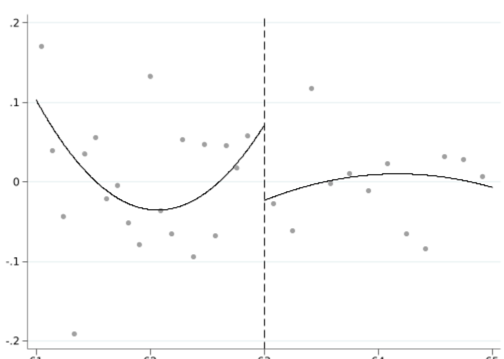


(a) RDD Plot

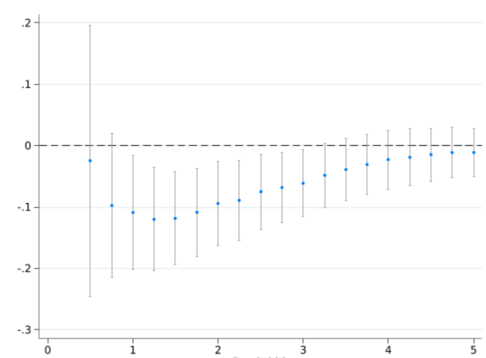


(b) Estimate when changing bandwidth

Figure A.44: Cutoff: 63; Outcome: Leaves School; Running Variable: School Lag; Sample: Top 25% *VA*

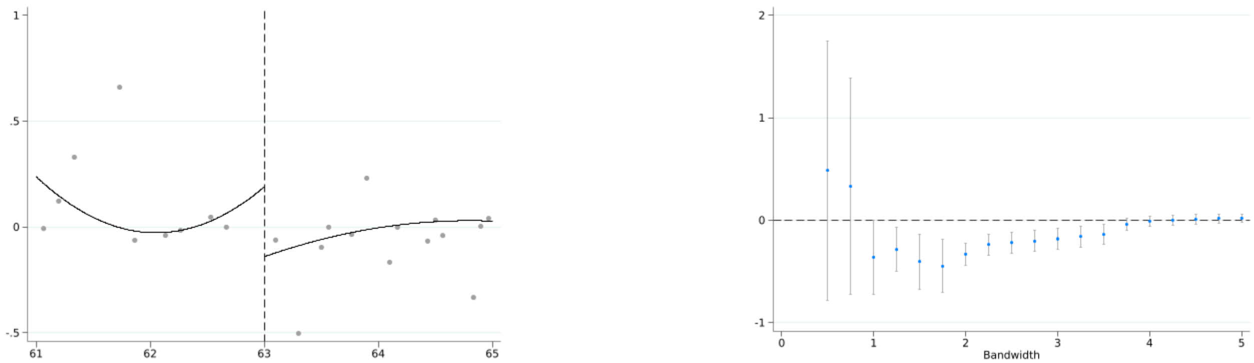


(a) RDD Plot



(b) Estimate when changing bandwidth

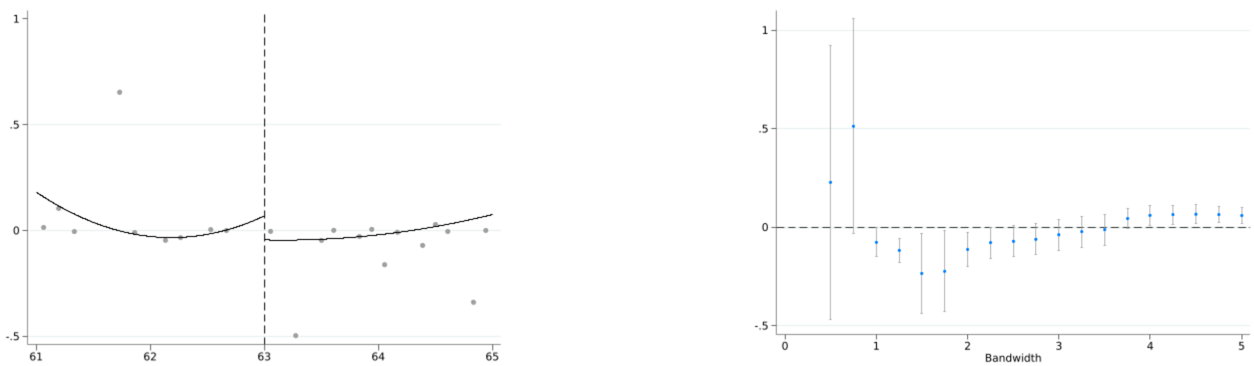
Figure A.45: Cutoff: 63; Outcome: Leaves WDPI; Running Variable: School Lag; Sample: Top 25% *VA*



(a) RDD Plot

(b) Estimate when changing bandwidth

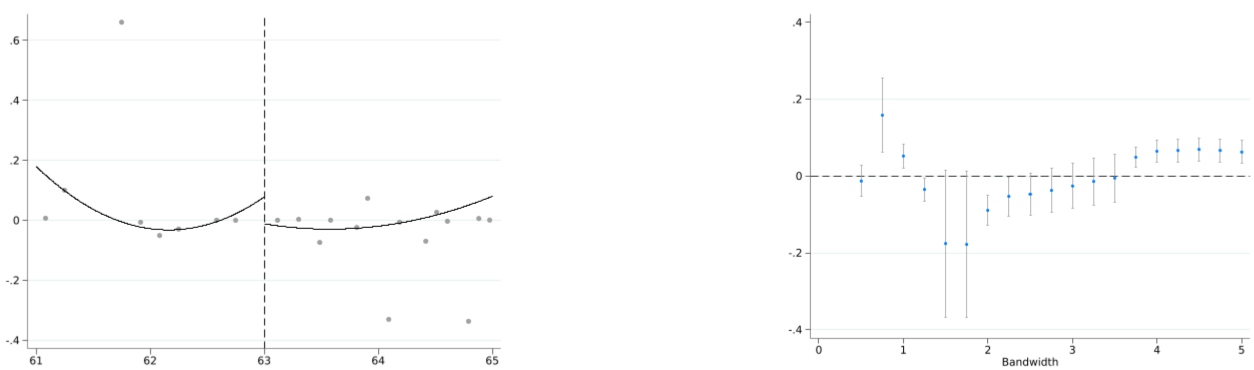
Figure A.46: Cutoff: 63; Outcome: Leaves School; Running Variable: District Lag; Sample: Top 25% *VA*



(a) RDD Plot

(b) Estimate when changing bandwidth

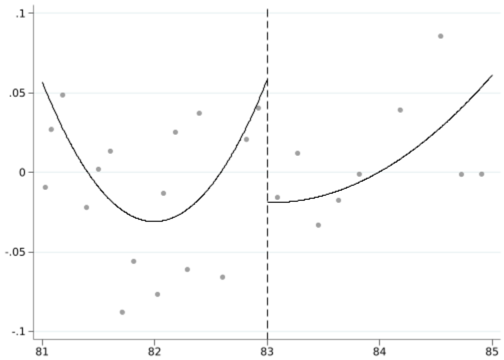
Figure A.47: Cutoff: 63; Outcome: Leaves District; Running Variable: District Lag; Sample: Top 25% *VA*



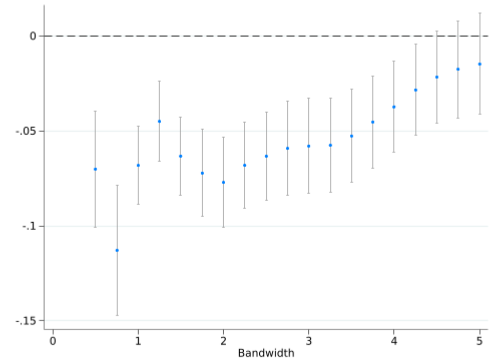
(a) RDD Plot

(b) Estimate when changing bandwidth

Figure A.48: Cutoff: 63; Outcome: Leaves WDPI; Running Variable: District Lag; Sample: Top 25% *VA*

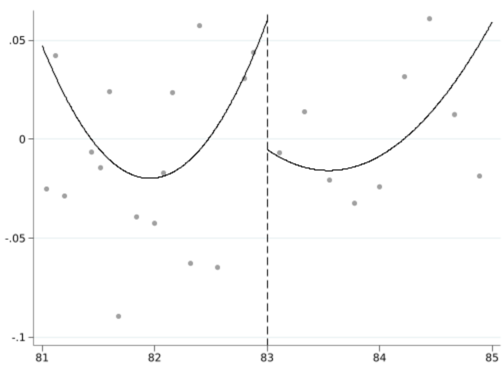


(a) RDD Plot

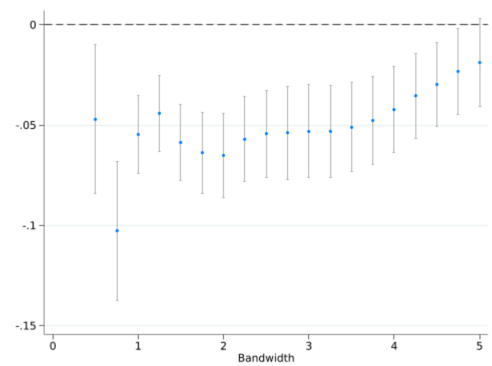


(b) Estimate when changing bandwidth

Figure A.49: Cutoff: 83; Outcome: Leaves School; Running Variable: District Lag; Sample: Top 25% *VA*

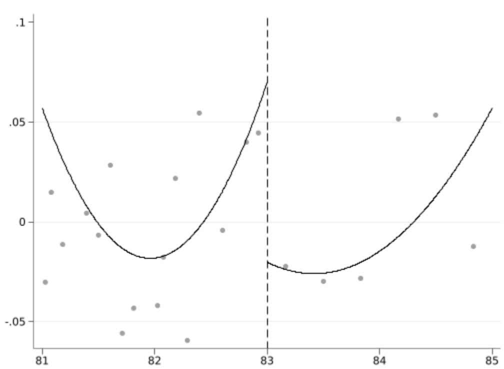


(a) RDD Plot

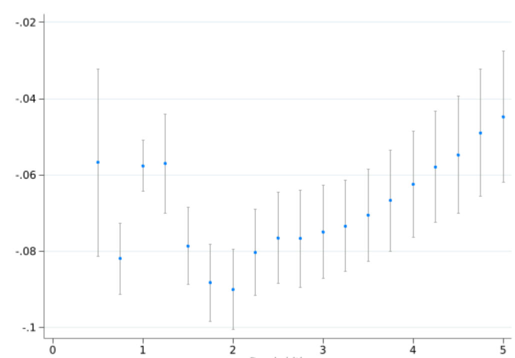


(b) Estimate when changing bandwidth

Figure A.50: Cutoff: 83; Outcome: Leaves District; Running Variable: District Lag; Sample: Top 25% *VA*



(a) RDD Plot



(b) Estimate when changing bandwidth

Figure A.51: Cutoff: 83; Outcome: Leaves WDPI; Running Variable: District Lag; Sample: Top 25% *VA*

A.2 Tables

Table A.1: Teacher Summary Statistics, School Accountability

	Total			Fails to Meet Expectations			Meets Few Expectations			Meets Expectations			Exceeds Expectations			Sig. Exceeds Expectations		
	All	Stayers	Leavers	All	Stayers	Leavers	All	Stayers	Leavers	All	Stayers	Leavers	All	Stayers	Leavers	All	Stayers	Leavers
Leaves current school	0.1581	0.0000	1.0000	0.3086	0.0000	1.0000	0.2448	0.0000	1.0000	0.1547	0.0000	1.0000	0.1366	0.0000	1.0000	0.1276	0.0000	1.0000
	(0.3648)	(0.0000)	(0.0000)	(0.4620)	(0.0000)	(0.0000)	(0.4300)	(0.0000)	(0.0000)	(0.3616)	(0.0000)	(0.0000)	(0.3435)	(0.0000)	(0.0000)	(0.3337)	(0.0000)	(0.0000)
Leaves to better rated school	0.0161	0.0000	0.1016	0.0573	0.0000	0.1858	0.0390	0.0000	0.1592	0.0179	0.0000	0.1158	0.0094	0.0000	0.0686	0.0031	0.0000	0.0240
	(0.1257)	(0.0000)	(0.3021)	(0.2325)	(0.0000)	(0.3891)	(0.1935)	(0.0000)	(0.3659)	(0.1327)	(0.0000)	(0.3201)	(0.0964)	(0.0000)	(0.2528)	(0.0552)	(0.0000)	(0.1531)
Leaves current district	0.1165	0.0000	0.7373	0.2114	0.0000	0.6851	0.1751	0.0000	0.7153	0.1148	0.0000	0.7420	0.1018	0.0000	0.7452	0.0971	0.0000	0.7610
	(0.3209)	(0.0000)	(0.4401)	(0.4084)	(0.0000)	(0.4647)	(0.3800)	(0.0000)	(0.4514)	(0.3188)	(0.0000)	(0.4376)	(0.3024)	(0.0000)	(0.4358)	(0.2961)	(0.0000)	(0.4266)
Leaves the WDPI	0.0885	0.0000	0.5600	0.1913	0.0000	0.6198	0.1494	0.0000	0.6105	0.0838	0.0000	0.5415	0.0748	0.0000	0.5472	0.0705	0.0000	0.5527
	(0.2840)	(0.0000)	(0.4964)	(0.3934)	(0.0000)	(0.4856)	(0.3565)	(0.0000)	(0.4877)	(0.2771)	(0.0000)	(0.4983)	(0.2630)	(0.0000)	(0.4978)	(0.2560)	(0.0000)	(0.4973)
Black	0.0205	0.0167	0.0402	0.2194	0.2081	0.2446	0.0832	0.0763	0.1044	0.0089	0.0076	0.0162	0.0039	0.0034	0.0076	0.0035	0.0032	0.0050
	(0.1415)	(0.1283)	(0.1964)	(0.4139)	(0.4060)	(0.4300)	(0.2761)	(0.2654)	(0.3059)	(0.0940)	(0.0868)	(0.1262)	(0.0626)	(0.0578)	(0.0870)	(0.0588)	(0.0569)	(0.0704)
Hispanic	0.0199	0.0181	0.0293	0.0410	0.0410	0.0409	0.0724	0.0708	0.0774	0.0200	0.0186	0.0277	0.0096	0.0090	0.0136	0.0056	0.0052	0.0086
	(0.1397)	(0.1335)	(0.1686)	(0.1982)	(0.1983)	(0.1981)	(0.2592)	(0.2565)	(0.2673)	(0.1401)	(0.1352)	(0.1641)	(0.0975)	(0.0943)	(0.1158)	(0.0746)	(0.0717)	(0.0924)
Asian	0.0082	0.0078	0.0106	0.0162	0.0163	0.0158	0.0168	0.0152	0.0218	0.0091	0.0090	0.0095	0.0052	0.0051	0.0064	0.0064	0.0060	0.0091
	(0.0904)	(0.0879)	(0.1026)	(0.1261)	(0.1268)	(0.1247)	(0.1286)	(0.1224)	(0.1459)	(0.0949)	(0.0945)	(0.0970)	(0.0722)	(0.0710)	(0.0796)	(0.0798)	(0.0774)	(0.0947)
Male	0.1265	0.1226	0.1472	0.1725	0.1611	0.1980	0.1416	0.1325	0.1697	0.1229	0.1197	0.1402	0.1235	0.1212	0.1382	0.1221	0.1200	0.1358
	(0.3324)	(0.3280)	(0.3544)	(0.3778)	(0.3676)	(0.3986)	(0.3486)	(0.3390)	(0.3754)	(0.3283)	(0.3246)	(0.3472)	(0.3290)	(0.3264)	(0.3452)	(0.3274)	(0.3250)	(0.3427)
Bachelor's Degree	0.5103	0.4985	0.5731	0.6832	0.6772	0.6966	0.5942	0.5732	0.6590	0.5074	0.4984	0.5563	0.4888	0.4805	0.5408	0.4752	0.4669	0.5319
	(0.4999)	(0.5000)	(0.4946)	(0.4653)	(0.4676)	(0.4599)	(0.4911)	(0.4946)	(0.4741)	(0.5000)	(0.5000)	(0.4969)	(0.4999)	(0.4996)	(0.4984)	(0.4994)	(0.4989)	(0.4991)
Master's or Higher	0.4892	0.5013	0.4249	0.3159	0.3215	0.3034	0.4054	0.4265	0.3401	0.4919	0.5013	0.4405	0.5108	0.5193	0.4574	0.5245	0.5330	0.4667
	(0.4999)	(0.5000)	(0.4943)	(0.4649)	(0.4671)	(0.4599)	(0.4910)	(0.4946)	(0.4738)	(0.4999)	(0.5000)	(0.4965)	(0.4999)	(0.4996)	(0.4982)	(0.4994)	(0.4989)	(0.4990)
Total Salary (2012 USD / 1000)	51.3796	51.7210	49.5607	52.1854	52.5127	51.4521	52.2217	52.6808	50.8049	50.6350	50.9196	49.0799	51.3783	51.7655	48.9317	52.6121	52.9597	50.2359
	(12.5984)	(12.3212)	(13.8415)	(13.6306)	(13.4428)	(14.0196)	(13.3494)	(13.1282)	(13.9169)	(12.3209)	(12.0348)	(13.6762)	(12.4663)	(12.2005)	(13.7815)	(12.7483)	(12.4969)	(14.1258)
Local Experience	11.8207	11.9870	10.9354	10.9996	11.1971	10.5572	11.4717	11.6763	10.8401	12.0119	12.1678	11.1600	11.8753	12.0261	10.9223	11.6003	11.7472	10.5956
	(8.7643)	(8.5964)	(9.5610)	(8.1217)	(7.9240)	(8.5344)	(8.2806)	(8.0980)	(8.7916)	(8.8953)	(8.7080)	(9.8132)	(8.8345)	(8.6620)	(9.8017)	(8.6760)	(8.5268)	(9.5761)
Total Experience	13.9409	14.0915	13.1386	11.8304	11.9776	11.5007	12.7458	12.9786	12.0274	14.0629	14.1764	13.4423	14.1617	14.2717	13.4665	14.3851	14.4865	13.6916
	(9.2121)	(8.9916)	(10.2709)	(8.4249)	(8.1901)	(8.9229)	(8.6641)	(8.4670)	(9.2104)	(9.3337)	(9.0957)	(10.5185)	(9.2752)	(9.0441)	(10.5943)	(9.1356)	(8.9315)	(10.4001)
FTE	0.9582	0.9649	0.9221	0.9675	0.9769	0.9463	0.9635	0.9714	0.9392	0.9589	0.9657	0.9220	0.9566	0.9632	0.9152	0.9543	0.9616	0.9045
	(0.1300)	(0.1192)	(0.1725)	(0.1238)	(0.1074)	(0.1523)	(0.1256)	(0.1119)	(0.1583)	(0.1281)	(0.1172)	(0.1713)	(0.1316)	(0.1215)	(0.1776)	(0.1353)	(0.1242)	(0.1872)
Observations	136081	114572	21509	4517	3123	1394	13144	9927	3217	48319	40843	7476	52789	45576	7213	17312	15103	2209

Notes: Standard deviations are reported in parentheses beneath the sample means.

Table A.2: Summary Statistics, District Accountability

	Total			Fails to Meet Expectations			Meets Few Expectations			Meets Expectations			Exceeds Expectations			Sig. Exceeds Expectations		
	All	Stayers	Leavers	All	Stayers	Leavers	All	Stayers	Leavers	All	Stayers	Leavers	All	Stayers	Leavers	All	Stayers	Leavers
Leaves current school	0.1581	0.0000	1.0000	0.2841	0.0000	1.0000	0.2661	0.0000	1.0000	0.1425	0.0000	1.0000	0.1327	0.0000	1.0000	0.1239	0.0000	1.0000
	(0.3648)	(0.0000)	(0.0000)	(0.4510)	(0.0000)	(0.0000)	(0.4419)	(0.0000)	(0.0000)	(0.3495)	(0.0000)	(0.0000)	(0.3392)	(0.0000)	(0.0000)	(0.3295)	(0.0000)	(0.0000)
Leaves current district	0.1165	0.0000	0.7373	0.2175	0.0000	0.7655	0.2132	0.0000	0.8012	0.1041	0.0000	0.7308	0.0978	0.0000	0.7375	0.0995	0.0000	0.8033
	(0.3209)	(0.0000)	(0.4401)	(0.4126)	(0.0000)	(0.4238)	(0.4096)	(0.0000)	(0.3992)	(0.3054)	(0.0000)	(0.4436)	(0.2971)	(0.0000)	(0.4400)	(0.2994)	(0.0000)	(0.3978)
Leaves to better rated district	0.0079	0.0000	0.0500	0.0080	0.0000	0.0282	0.0198	0.0000	0.0742	0.0091	0.0000	0.0637	0.0040	0.0000	0.0305	0.0020	0.0000	0.0163
	(0.0886)	(0.0000)	(0.2180)	(0.0892)	(0.0000)	(0.1657)	(0.1392)	(0.0000)	(0.2622)	(0.0948)	(0.0000)	(0.2443)	(0.0635)	(0.0000)	(0.1719)	(0.0449)	(0.0000)	(0.1266)
Leaves the WDPI	0.0885	0.0000	0.5600	0.2030	0.0000	0.7144	0.1843	0.0000	0.6924	0.0743	0.0000	0.5217	0.0687	0.0000	0.5180	0.0662	0.0000	0.5346
	(0.2840)	(0.0000)	(0.4964)	(0.4023)	(0.0000)	(0.4518)	(0.3877)	(0.0000)	(0.4616)	(0.2623)	(0.0000)	(0.4996)	(0.2530)	(0.0000)	(0.4997)	(0.2487)	(0.0000)	(0.4991)
Black	0.0205	0.0167	0.0402	0.1566	0.1519	0.1683	0.0790	0.0625	0.1246	0.0050	0.0050	0.0050	0.0024	0.0022	0.0040	0.0034	0.0035	0.0027
	(0.1415)	(0.1283)	(0.1964)	(0.3634)	(0.3590)	(0.3742)	(0.2698)	(0.2421)	(0.3303)	(0.0706)	(0.0706)	(0.0706)	(0.0492)	(0.0467)	(0.0631)	(0.0579)	(0.0587)	(0.0521)
Hispanic	0.0199	0.0181	0.0293	0.0899	0.0907	0.0879	0.0693	0.0704	0.0665	0.0144	0.0136	0.0192	0.0048	0.0043	0.0080	0.0045	0.0044	0.0054
	(0.1397)	(0.1335)	(0.1686)	(0.2861)	(0.2872)	(0.2833)	(0.2540)	(0.2558)	(0.2492)	(0.1193)	(0.1160)	(0.1372)	(0.0693)	(0.0657)	(0.0890)	(0.0672)	(0.0663)	(0.0735)
Asian	0.0082	0.0078	0.0106	0.0201	0.0203	0.0195	0.0138	0.0120	0.0187	0.0086	0.0084	0.0096	0.0048	0.0045	0.0067	0.0035	0.0036	0.0027
	(0.0904)	(0.0879)	(0.1026)	(0.1402)	(0.1409)	(0.1385)	(0.1166)	(0.1089)	(0.1356)	(0.0922)	(0.0913)	(0.0975)	(0.0693)	(0.0672)	(0.0817)	(0.0593)	(0.0603)	(0.0521)
Male	0.1265	0.1226	0.1472	0.1717	0.1642	0.1906	0.1387	0.1238	0.1797	0.1195	0.1168	0.1356	0.1244	0.1222	0.1387	0.1320	0.1297	0.1479
	(0.3324)	(0.3280)	(0.3544)	(0.3771)	(0.3705)	(0.3928)	(0.3456)	(0.3294)	(0.3840)	(0.3243)	(0.3212)	(0.3424)	(0.3300)	(0.3275)	(0.3457)	(0.3385)	(0.3360)	(0.3552)
Bachelor's Degree	0.5103	0.4985	0.5731	0.6745	0.6688	0.6889	0.6160	0.5657	0.7549	0.5024	0.4964	0.5385	0.4914	0.4841	0.5388	0.4677	0.4611	0.5142
	(0.4999)	(0.5000)	(0.4946)	(0.4686)	(0.4707)	(0.4631)	(0.4864)	(0.4957)	(0.4302)	(0.5000)	(0.5000)	(0.4985)	(0.4999)	(0.4998)	(0.4985)	(0.4990)	(0.4985)	(0.5001)
Master's or Higher	0.4892	0.5013	0.4249	0.3252	0.3307	0.3111	0.3831	0.4332	0.2448	0.4968	0.5034	0.4570	0.5085	0.5157	0.4610	0.5323	0.5389	0.4858
	(0.4999)	(0.5000)	(0.4943)	(0.4685)	(0.4705)	(0.4631)	(0.4862)	(0.4956)	(0.4300)	(0.5000)	(0.5000)	(0.4982)	(0.4999)	(0.4998)	(0.4985)	(0.4990)	(0.4985)	(0.5001)
Total Salary (2012 USD / 1000)	51.3796	51.7210	49.5607	49.4459	51.1080	45.2582	53.7399	53.6801	53.9045	50.4727	50.8370	48.2795	50.8076	51.1676	48.4538	54.3386	54.7811	51.2099
	(12.5984)	(12.3212)	(13.8415)	(14.3211)	(14.5871)	(12.7009)	(12.0410)	(11.6710)	(13.0079)	(12.2605)	(12.0122)	(13.4539)	(12.2195)	(11.9861)	(13.4128)	(13.2582)	(13.0229)	(14.4430)
Local Experience	11.8207	11.9870	10.9354	11.8367	12.5142	10.1298	12.4508	12.4499	12.4535	11.8326	12.0490	10.5300	11.4576	11.6434	10.2427	11.4528	11.5793	10.5590
	(8.7643)	(8.5964)	(9.5610)	(7.9808)	(7.7339)	(8.3328)	(8.9348)	(8.7205)	(9.5024)	(8.9743)	(8.8257)	(9.7205)	(8.7532)	(8.6182)	(9.5011)	(8.6697)	(8.5327)	(9.5413)
Total Experience	13.9409	14.0915	13.1386	12.4219	13.0659	10.7992	13.9377	14.0050	13.7521	13.9601	14.1242	12.9720	13.9946	14.1206	13.1710	14.7726	14.8474	14.2436
	(9.2121)	(8.9916)	(10.2709)	(8.1917)	(7.8901)	(8.6997)	(9.3304)	(9.1355)	(9.8473)	(9.3947)	(9.1849)	(10.5165)	(9.2022)	(9.0105)	(10.3321)	(9.1487)	(8.9555)	(10.4041)
FTE	0.9582	0.9649	0.9221	0.9704	0.9805	0.9449	0.9700	0.9733	0.9609	0.9570	0.9640	0.9149	0.9560	0.9628	0.9119	0.9574	0.9628	0.9192
	(0.1300)	(0.1192)	(0.1725)	(0.1171)	(0.0962)	(0.1550)	(0.1179)	(0.1140)	(0.1278)	(0.1305)	(0.1195)	(0.1777)	(0.1324)	(0.1221)	(0.1803)	(0.1306)	(0.1226)	(0.1728)
Observations	136081	114572	21509	6483	4641	1842	10225	7504	2721	50455	43267	7188	35872	31113	4759	5948	5211	737

Notes: Standard deviations are reported in parentheses beneath the sample means.

Table A.3: Turnover RDD Estimates; Running Variable: School Accountability Score; Sample: Bottom 25% *VA*

	Cutoff: 53		Cutoff: 63		Cutoff: 73		Cutoff: 83	
	Leaves current school	Leaves the WDPI	Leaves current school	Leaves the WDPI	Leaves current school	Leaves the WDPI	Leaves current school	Leaves the WDPI
Accountability Score (Lag)	0.0192	0.0268	0.1936***	0.0445	-0.0174	0.0309	-0.2419**	-0.0540
	(0.1084)	(0.0424)	(0.0567)	(0.0520)	(0.0395)	(0.0297)	(0.0889)	(0.0601)
Observations	686	686	3066	3066	5102	5102	1276	1276
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	District	District	District	District	District	District	District	District
Bandwidth	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000
Kernel	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.4: Turnover RDD Estimates; Running Variable: District Accountability Score; Sample: Bottom 25% *VA*

	Cutoff: 63			Cutoff: 73			Cutoff: 83		
	Leaves current school	Leaves current district	Leaves the WDPI	Leaves current school	Leaves current district	Leaves the WDPI	Leaves current school	Leaves current district	Leaves the WDPI
Accountability Score (Lag)	-0.1613**	-0.1144*	-0.0674*	0.0856	0.0677	0.0530	0.1767*	0.1673***	0.0858*
	(0.0507)	(0.0458)	(0.0336)	(0.0511)	(0.0407)	(0.0297)	(0.0797)	(0.0489)	(0.0369)
Observations	2723	2723	2723	4395	4395	4395	734	734	734
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	District	District	District	District	District	District	District	District	District
Bandwidth	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000
Kernel	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.5: Turnover RDD Estimates; Running Variable: School Accountability Score; Sample: Top 25% *VA*

	Cutoff: 53		Cutoff: 63		Cutoff: 73		Cutoff: 83	
	Leaves current school	Leaves the WDPI	Leaves current school	Leaves the WDPI	Leaves current school	Leaves the WDPI	Leaves current school	Leaves the WDPI
Accountability Score (Lag)	-0.1108	0.1188***	-0.1233**	-0.0943**	0.0249	0.0391	0.0064	0.0075
	(0.1830)	(0.0248)	(0.0400)	(0.0352)	(0.0333)	(0.0242)	(0.0282)	(0.0223)
Observations	1157	1157	2845	2845	8067	8067	7316	7316
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	District	District	District	District	District	District	District	District
Bandwidth	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000
Kernel	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.6: Turnover RDD Estimates; Running Variable: District Accountability Score; Sample: Top 25% *VA*

	Cutoff: 63			Cutoff: 73			Cutoff: 83		
	Leaves current school	Leaves current district	Leaves the WDPI	Leaves current school	Leaves current district	Leaves the WDPI	Leaves current school	Leaves current district	Leaves the WDPI
Accountability Score (Lag)	-0.3299***	-0.1123*	-0.0894***	0.0026	0.0216	0.0289	-0.0772***	-0.0650***	-0.0901***
	(0.0557)	(0.0437)	(0.0198)	(0.0500)	(0.0322)	(0.0198)	(0.0121)	(0.0107)	(0.0054)
Observations	2961	2961	2961	7696	7696	7696	4549	4549	4549
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	District	District	District	District	District	District	District	District	District
Bandwidth	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000
Kernel	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.7: Turnover RDD Estimates; Running Variable: School Accountability Score; Sample: Bottom 25% *VA* (No Milwaukee)

	Cutoff: 53		Cutoff: 63		Cutoff: 73		Cutoff: 83	
	Leaves current school	Leaves the WDPI	Leaves current school	Leaves the WDPI	Leaves current school	Leaves the WDPI	Leaves current school	Leaves the WDPI
Accountability Score (Lag)	-0.1274**	-0.0127	0.1967**	0.0792*	-0.0222	0.0247	-0.2261*	-0.0081
	(0.0481)	(0.0133)	(0.0639)	(0.0362)	(0.0431)	(0.0315)	(0.1079)	(0.0368)
Observations	332	332	2485	2485	4654	4654	1109	1109
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	District	District	District	District	District	District	District	District
Bandwidth	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000
Kernel	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.8: Turnover RDD Estimates; Running Variable: District Accountability Score; Sample: Bottom 25% *VA* (No Milwaukee)

	Cutoff: 63			Cutoff: 73			Cutoff: 83		
	Leaves current school	Leaves current district	Leaves the WDPI	Leaves current school	Leaves current district	Leaves the WDPI	Leaves current school	Leaves current district	Leaves the WDPI
Accountability Score (Lag)	-0.2373***	-0.1627**	-0.0925*	0.0777	0.0615	0.0474	0.0426	-0.1032	-0.1840***
	(0.0567)	(0.0500)	(0.0436)	(0.0517)	(0.0409)	(0.0299)	(0.1367)	(0.0884)	(0.0419)
Observations	2353	2353	2353	4103	4103	4103	634	634	634
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	District	District	District	District	District	District	District	District	District
Bandwidth	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000
Kernel	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.9: Turnover RDD Estimates; Running Variable: School Accountability Score; Sample: Top 25% *VA* (No Milwaukee)

	Cutoff: 53		Cutoff: 63		Cutoff: 73	
	Leaves current school	Leaves the WDPI	Leaves current school	Leaves the WDPI	Leaves current school	Leaves the WDPI
Accountability Score (Lag)	-0.0827	-0.1448	0.0396	0.0638**	-0.0063	0.0071
	(0.1533)	(0.1599)	(0.0402)	(0.0227)	(0.0313)	(0.0254)
Observations	1526	1526	5746	5746	4435	4435
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	District	District	District	District	District	District
Bandwidth	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000
Kernel	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.10: Turnover RDD Estimates; Running Variable: District Accountability Score; Sample: Top 25% *VA* (No Milwaukee)

	Cutoff: 63			Cutoff: 73			Cutoff: 83		
	Leaves current school	Leaves current district	Leaves the WDPI	Leaves current school	Leaves current district	Leaves the WDPI	Leaves current school	Leaves current district	Leaves the WDPI
Accountability Score (Lag)	-0.6975*	-0.4906	-0.0759	-0.0027	0.0200	0.0306	-0.0016	0.0432	-0.0699***
	(0.3372)	(0.3309)	(0.0913)	(0.0506)	(0.0320)	(0.0209)	(0.0850)	(0.0748)	(0.0125)
Observations	2090	2090	2090	5971	5971	5971	2258	2258	2258
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	District	District	District	District	District	District	District	District	District
Bandwidth	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000
Kernel	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$