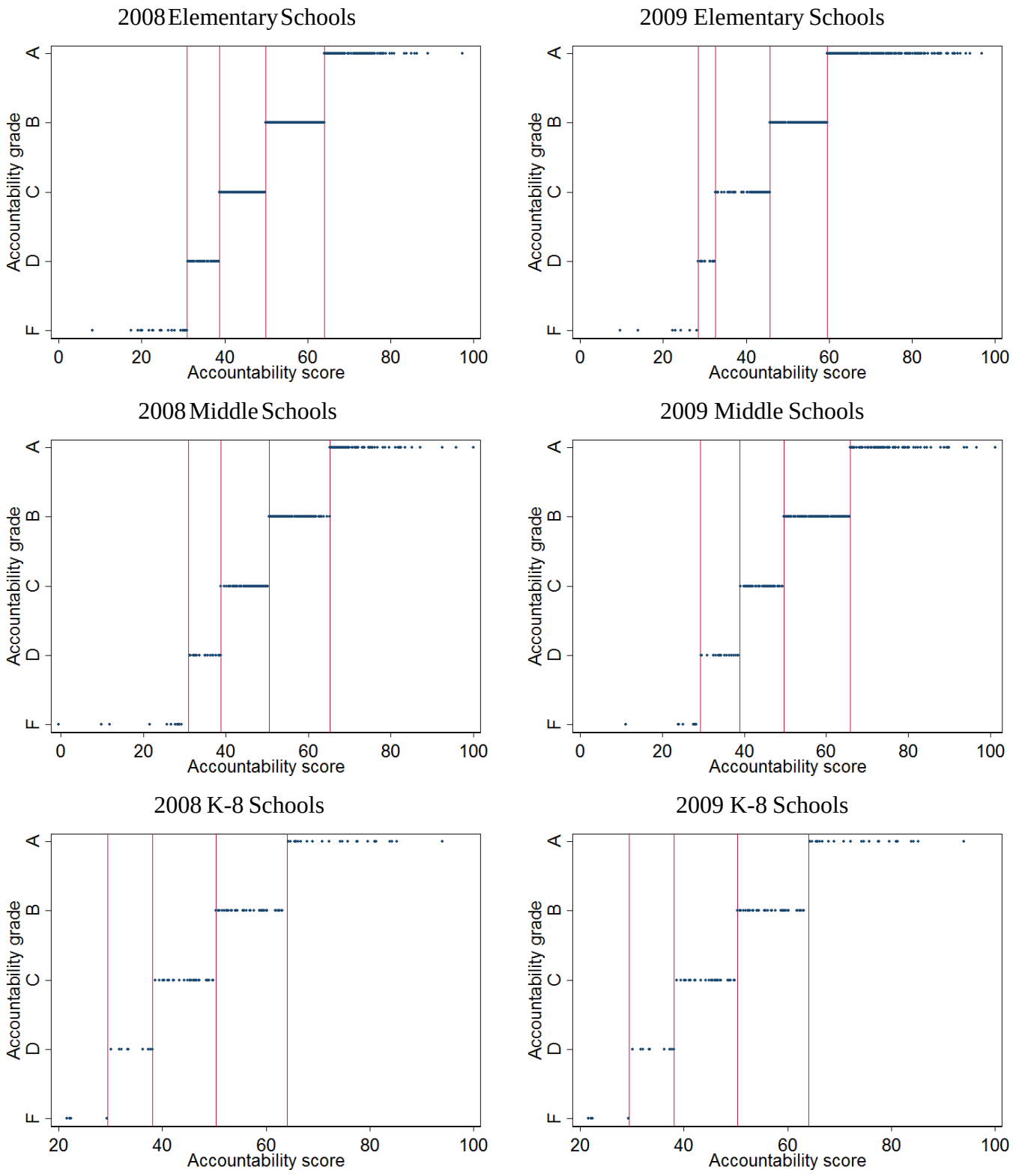
How Does School Accountability Affect Teachers? Evidence from New York City

Rebecca Dizon-Ross

Online Appendix

Appendix Figure A1: Relationship between Accountability Scores and Accountability Grades

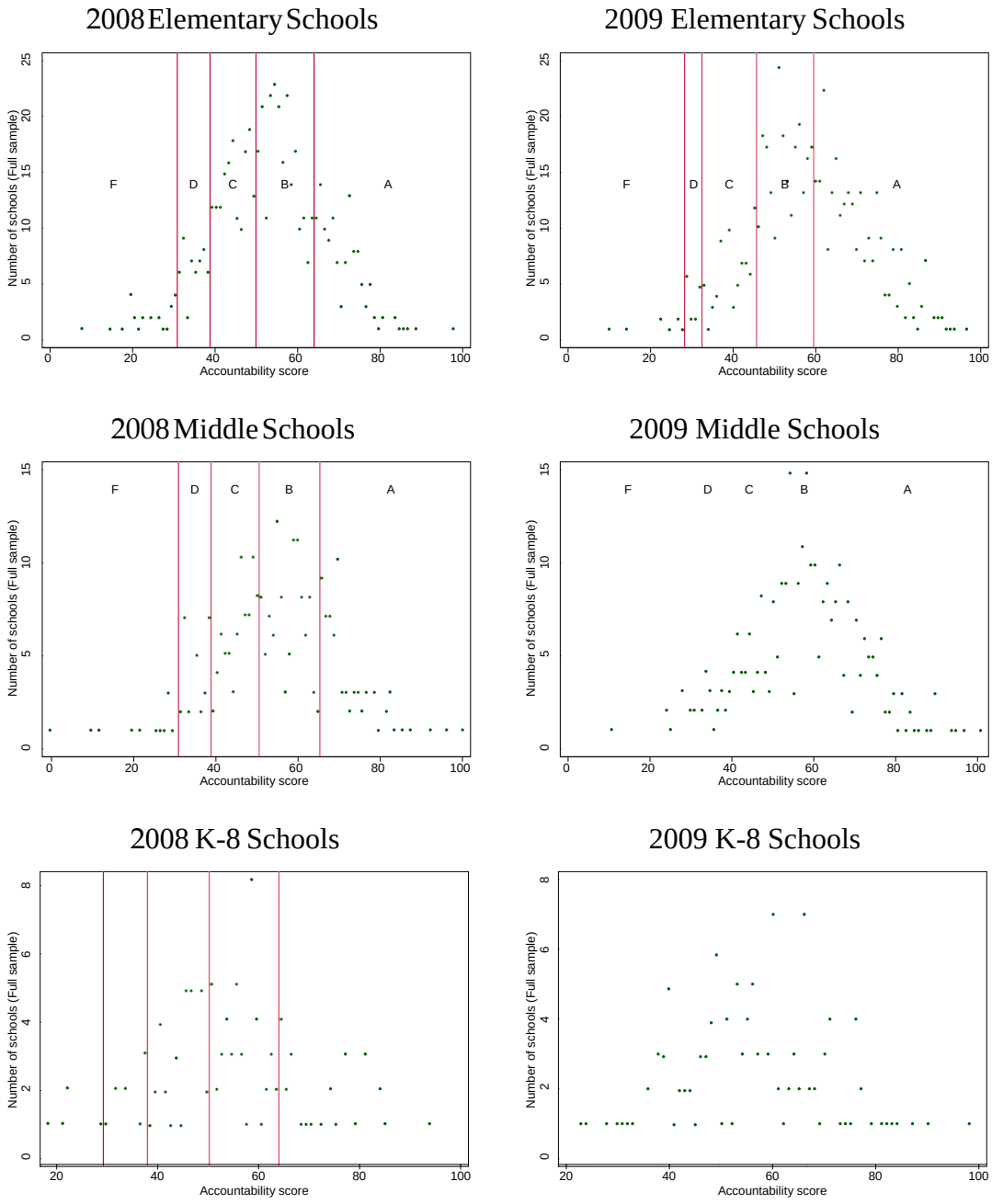
(Each year and schooltype shown separately)



Notes. For each year and school type, the figures plot the accountability grade received by a school as a function of the underlying accountability score.

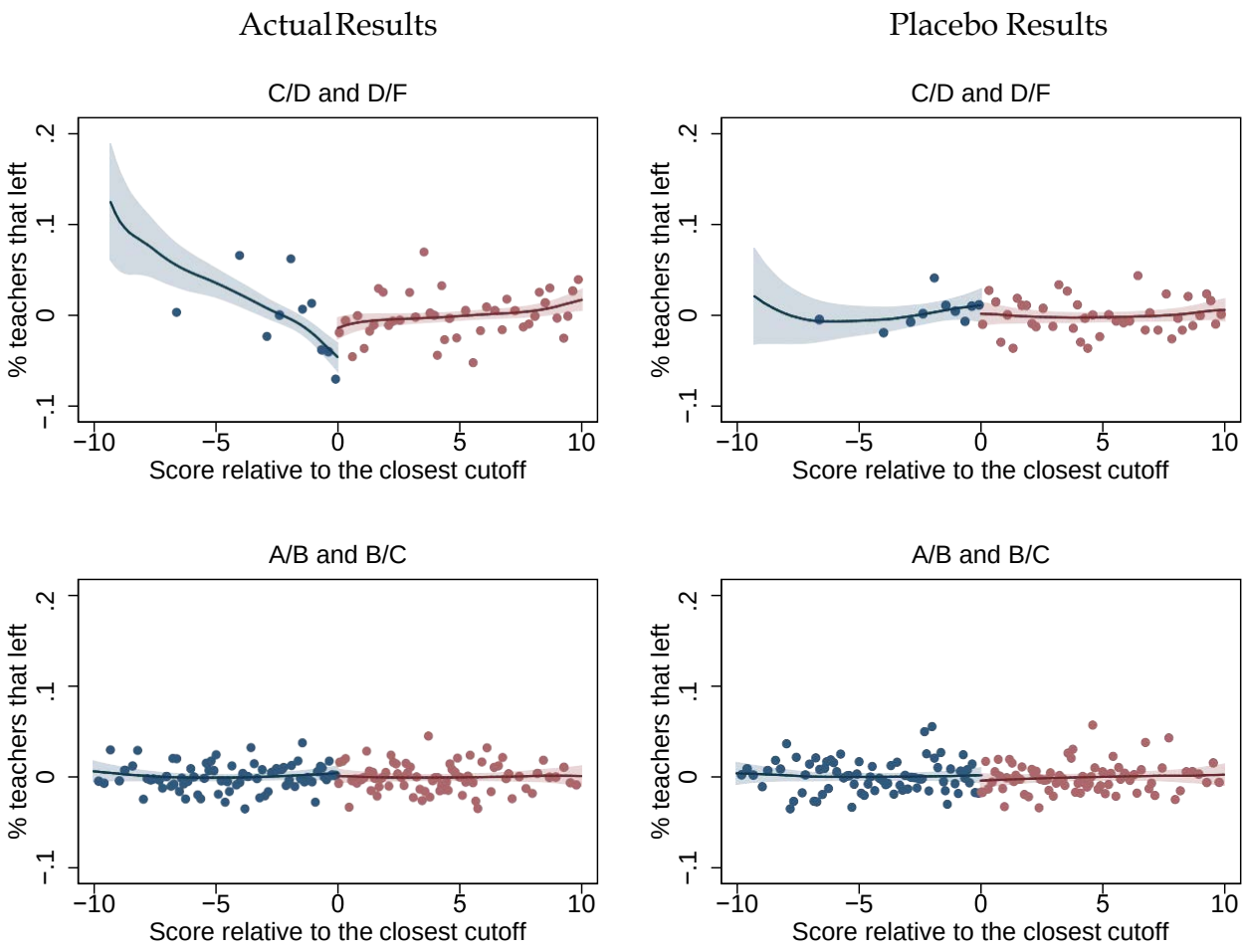
Appendix Figure A2: Density of Schools Near Grade Thresholds

(Each year and schooltype shown separately)



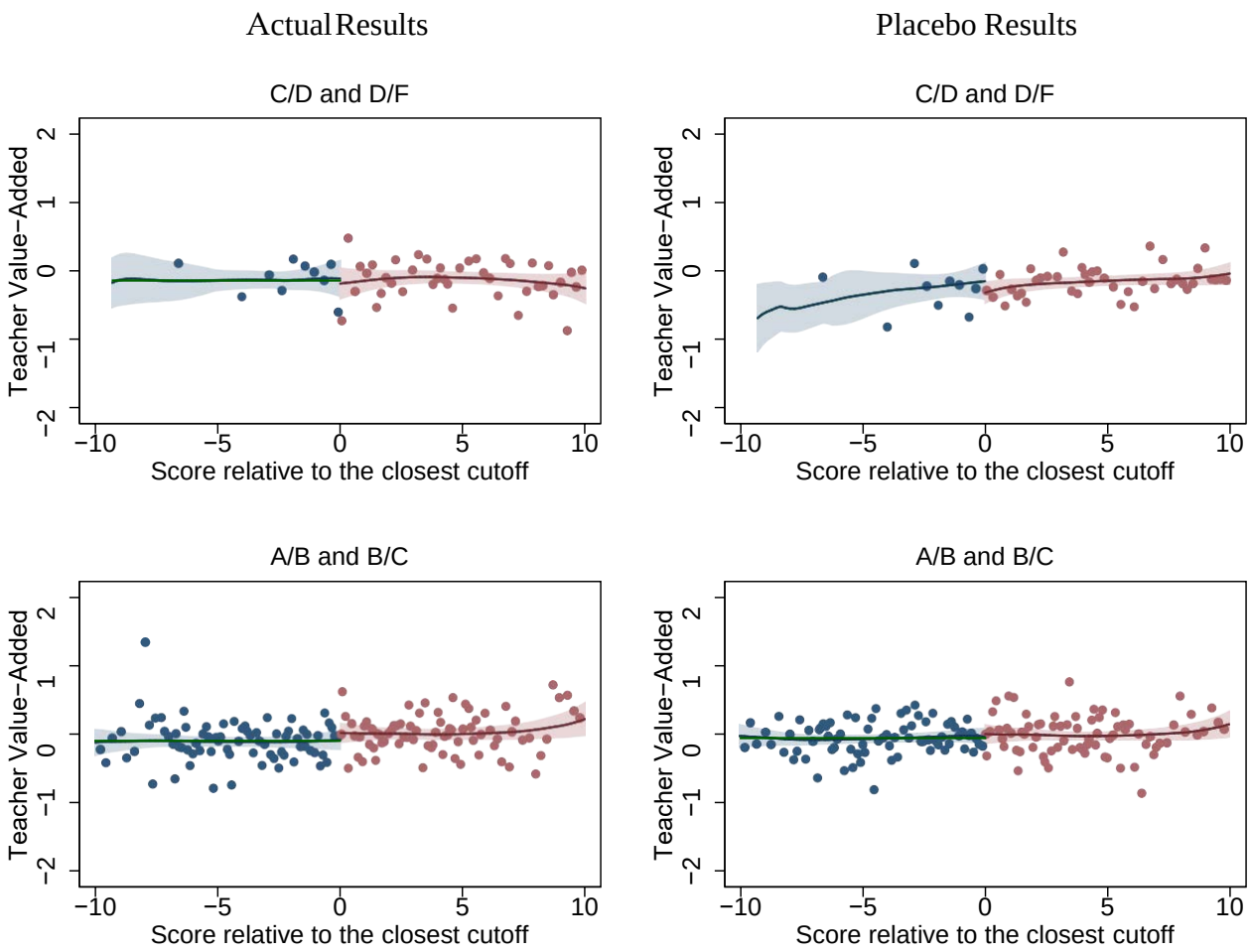
Notes. For each year and school type, the figures plot the number of schools with a given accountability score (specifically, the y-axis shows the number of schools within a 0.5 point bandwidth of the accountability score displayed on the X-axis). The red lines show the 4 grade thresholds (A/B, B/C, C/D, and D/F). Evidence of heaping directly adjacent to the grade thresholds line would be a violation of the regression discontinuity identification assumptions.

Appendix Figure A3: Residual Turnover, by Accountability Score
(version with a fixed number of schools per dot)



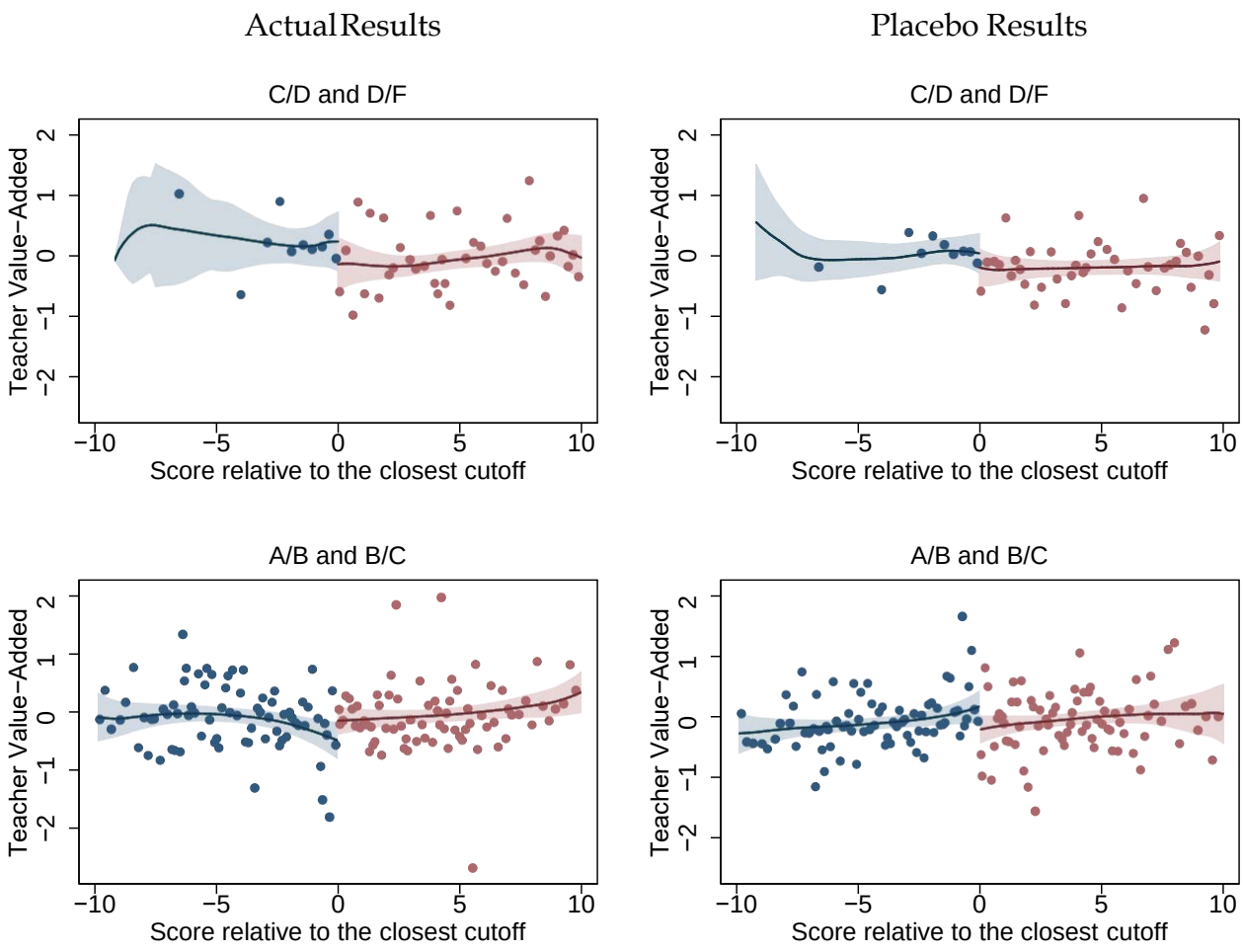
Notes. Each dot represents 10 schools. The left column plots the actual turnover results. The x-axes show the schools' average accountability scores relative to the closest grade threshold (so the grade threshold is always displayed at 0) and the y-axes show average residual turnover in the summer after the schools received their accountability grades. The right column shows placebo turnover results: there, the y-axes show residual turnover in the year *before* schools received grades (placebo uses data from the first year only). Residual turnover is calculated by regressing an indicator for leaving a school on a vector of covariates (see Table 2 notes for list of covariates). The lines correspond to local polynomial smooth plots, and the shaded areas represent their 95% confidence intervals.

Appendix Figure A4: Average Math Value-Added of Leavers, by Accountability Score
(version with a fixed number of schools per dot)



Notes. Each dot represents 10 schools. The left column plots the actual leaver quality results. The x-axes show the schools' average accountability scores relative to the closest grade threshold (so the grade threshold is always displayed at 0). The y-axes show the average value-added of leavers (i.e., of the teachers who left their schools in the summer after their schools received the accountability score and grade). The right panel has the placebo results: there, the y-axes show the average value-added of the teachers who left their schools the year *before* their schools received the accountability score and grade (placebo uses data from the first year only). The lines correspond to local polynomial smooth plots, and the shaded areas represent their 95% confidence intervals.

Appendix Figure A5: Average Math Value-Added of Joiners, by Accountability Score
(version with a fixed number of schools per dot)



Notes. Each dot represents 10 schools. The left panel plots the actual joiner quality results. The x-axes show the schools' average accountability scores relative to the closest grade threshold (so the grade threshold is always displayed at 0). The y-axes show the average value-added of joiners (i.e., of the teachers who joined schools in the summer after their schools received the accountability score and grade). The right panel has the placebo results: there, the y-axes show the average value-added of the teachers who joined their schools the year *before* their schools received the accountability score and grade (placebo uses data from the first year only). The lines correspond to local polynomial smooth plots, and the shaded areas represent their 95% confidence intervals.

Appendix Table A1. McCrary test

Dependent Variable:	Number of schools in a bin	
Binwidth:	0.5	0.1
Independent Variable:	(1)	(2)
Bottom of the grade distribution (C/D and D/F)		
School received lower grade (dummy)	-2.585	-0.561
	[3.326]	[0.845]
N	20	100
Top of the grade distribution (A/B and B/C)		
School received lower grade (dummy)	-3.436	-0.807
	[4.865]	[1.281]
N	20	100

Notes. Table shows the results of the McCrary test for gaming. The dependent variable is the number of schools in a bin, using a binwidth of 0.5 (column (1)) or 0.1 point (column (2)). Each column contains two cells containing the results of two separate regressions (one with schools at the bottom of the grade distribution, one with those at the top). Regressions use a bandwidth of 5 grade points. All regressions include a control for the score relative to the closest threshold, allowed to vary on either side of the threshold. Robust standard errors are reported in brackets. Data come from the 2007-08 and 2008-09 school years in the New York City Department of Education, using the report card grade that was received by the school during fall of the school year. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix Table A2. No Selection Into the Analysis Sample

Dependent Variable:	School closed	School is missing teacher data	School being phased out	School not restructuring at baseline	School was part of New York School Bonus Program (NYSBP)
*Independent Variable = School received **lower** grade at the:*	(1)	(2)	(3)	(4)	(5)
Bottom of the grade distribution					
D/F or C/D thresholds (grouped)	0.000	-0.023	0.011	-0.107	0.000
	[0.000]	[0.020]	[0.032]	[0.080]	[0.000]
N	312	312	312	312	312
Dep. Var. Mean	0.003	0.006	0.038	0.901	0.173
Top of the grade distribution					
B/C or A/B Thresholds (grouped)	n/a	n/a	0.007	-0.014	0.000
			[0.005]	[0.031]	[0.000]
N	932	932	932	932	932
Dep. Var. Mean	0.000	0.000	0.002	0.905	0.157

Notes. The table presents regression discontinuity estimates of the effect of school accountability grades on school characteristics. Regressions use a bandwidth of 5 grade points. Standard errors are reported in brackets and clustered at the school level. Each observation is a school in a given year. Cols (1) and (2): The sample is all schools receiving accountability grades, and the dependent variable is an indicator that the school is excluded from the analysis sample because it closed or was missing teacher data (4 schools total). Cols (3)-(6): The sample is all non-excluded schools receiving accountability grades. Cols (7) and (8): The sample is all non-excluded schools that had not started restructuring prior to accountability. School controls include controls for the average previous year's achievement; the previous year's accountability score (second year only); the percent of students that are black, Hispanic, Asian, that receive free and reduced price lunch, and that are immigrants; whether the school was in the NYSBP; fixed effects for school size; and five-year average school turnover prior to the institution of accountability. All regressions control for the "RD running variable" (accountability score relative to the threshold) separately by (year)X(schooltype)X(received lower grade), as well as controls for (year)X(schooltype)X(received lower grade) and all of the main effects and lower-order interactions (e.g., year, schooltype, yearXschooltype, etc.). Data come from the 2007-08 and 2008-09 school years in the New York City Department of Education, using the report card grade that was received by the school during fall of the school year. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix Table A3. Covariate Balance

Panel A. Teacher characteristics

	Teacher characteristics												*Turnover*
Dependent Variable:	Math value-added (VA)	ELA VA	ELA predicted VA	Math predicted VA	Female	White	Asian	Black	Hispanic	Years of experience	Has master's degree	Age	Prior turnover
Independent Variable =	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
*School received **lower** grade at the:*													
Bottom of the grade distribution													
D/F or C/D thresholds (grouped)	0.050	0.012	0.001	0.000	0.009	0.067	-0.006	-0.088	0.028	-0.227	0.002	-0.594	0.008
	[0.10]	[0.08]	[0.00]	[0.00]	[0.02]	[0.06]	[0.01]	[0.05]*	[0.03]	[0.59]	[0.03]	[0.81]	[0.01]
N	306	306	309	309	309	309	309	309	309	309	309	309	309
Dep. Var. Mean	-0.13	-0.09	0.04	0.01	0.81	0.55	0.04	0.27	0.15	9.28	0.41	40.90	0.17
Top of the grade distribution													
B/C or A/B Thresholds	-0.007	-0.003	0.001	0.001	-0.005	-0.046	0.001	0.007	0.039	-0.290	-0.017	-0.069	0.015
	[0.053]	[0.044]	[0.000]	[0.000]*	[0.009]	[0.031]	[0.010]	[0.027]	[0.017]**	[0.310]	[0.017]	[0.457]	[0.008]*
N	927	928	932	932	932	932	932	932	932	932	932	932	932
Dep. Var. Mean	0.00	0.00	0.04	0.01	0.83	0.61	0.05	0.21	0.14	9.72	0.43	40.97	0.157

Panel B. Student and school characteristics

	Student body characteristics									*School characteristics*			
Dependent Variable:	Previous year math scores	Previous year ELA scores	Female	White	Asian	Black	Hispanic	Immigrants	Poverty status	ln(enrollment)	ln(# teachers)	New hires	Teacher bonus program
Independent Variable =	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
School received lower grade at the:													
Bottom of the grade distribution													
D/F or C/D thresholds (grouped)	0.016	-0.030	0.004	0.065	-0.016	-0.185	0.135	0.000	0.011	-0.145	-0.133	-0.194	0.049
	[0.08]	[0.08]	[0.01]	[0.05]	[0.02]	[0.06]***	[0.06]**	[0.00]	[0.05]	[0.12]	[0.11]	[1.12]	[0.10]
N	309	309	309	309	309	309	309	309	309	309	309	309	309
Dep. Var. Mean	-0.25	-0.18	0.49	0.12	0.07	0.42	0.39	0.02	0.71	6.33	3.91	7.78	0.17
Top of the grade distribution													
B/C or A/B Thresholds	-0.051	-0.056	0.003	-0.031	-0.031	0.013	0.050	0.003	0.026	0.088	0.041	0.361	-0.012
	[0.059]	[0.059]	[0.005]	[0.031]	[0.024]	[0.038]	[0.035]	[0.003]	[0.031]	[0.065]	[0.059]	[0.637]	[0.046]
N	932	932	932	932	932	932	932	932	932	932	932	932	932
Dep. Var. Mean	0.03	0.03	0.49	0.15	0.13	0.32	0.39	0.02	0.67	6.43	3.98	6.75	0.16

Notes. This table presents balance tests: regression discontinuity estimates where the outcome variables are teacher (Panel A), student (Panel B, columns (1)-(9)), or school characteristics (Panel B, columns (10)-(13)). Each observation is a school in a given year. All regressions include controls for the "RD running variable" (specifically, a control for (year)X(schooltype)X(received lower grade)) and all of the main effects and lower-order interactions (e.g., year, schooltype, yearXschooltype, etc.), and a dummy for receiving the lower accountability grade at the threshold). Poverty is the percent of students who receive free and reduced price lunch. Prior turnover is the control used in the regressions, which represents 5-year average turnover before the institution of accountability. Standard errors are reported in brackets. Data come from the 2008-09 and 2009-10 school years. All data from the New York City Department of Education. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix Table A4. Joiner and leaver characteristics: Full set of placebo tests

Sample:	Leavers				Joiners			
Independent Var. =	Estimated math VA	Predicted math VA	>=4 years experience	Has a masters	Estimated math VA	Predicted math VA	>=4 years experience	Has a masters
*Received **lower** grade at:*	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Both years of accountability system								
Bottom of the grade dist.								
D/F and C/D thresholds	0.107	-0.095	-0.033	-0.073	0.241	-0.087	0.036	0.047
	[0.17]	[0.09]	[0.05]	[0.04]*	[0.26]	[0.09]	[0.04]	[0.03]
N	496	2,118	2,118	2,118	227	2,404	2,404	2,404
Dep. Var. Mean	-0.16	-0.07	0.52	0.34	-0.11	-0.14	0.24	0.18
Top of the grade dist.								
A/B and B/C thresholds	0.016	-0.032	0.012	-0.006	0.324	0.069	-0.039	-0.013
	[0.10]	[0.05]	[0.03]	[0.03]	[0.17]*	[0.06]	[0.02]	[0.02]
N	1,274	5,628	5,628	5,628	521	6,292	6,292	6,292
Dep. Var. Mean	-0.02	-0.02	0.54	0.36	-0.06	-0.09	0.21	0.17
Panel B. First year of accountability system only								
Bottom of the grade dist.								
D/F and C/D thresholds	0.119	0.093	-0.032	-0.118	0.005	0.024	-0.001	0.013
	[0.21]	[0.13]	[0.08]	[0.05]**	[0.31]	[0.10]	[0.05]	[0.05]
N	328	1,392	1,392	1,392	163	1,655	1,655	1,655
Dep. Var. Mean	-0.17	0.00	0.55	0.36	-0.14	-0.11	0.24	0.18
Top of the grade dist.	-0.070	-0.019	0.021	-0.056	-0.036	0.040	-0.030	-0.054
A/B and B/C thresholds	[0.13]	[0.07]	[0.04]	[0.04]	[0.22]	[0.07]	[0.03]	[0.03]**
N	754	3,190	3,190	3,190	334	3,836	3,836	3,836
Dep. Var. Mean	-0.04	-0.03	0.54	0.36	-0.06	-0.11	0.21	0.18

Notes. Table presents regression discontinuity estimates of the effect of school accountability grades on the characteristics of the teachers that leave (leavers) and teachers that are hired by a given school (joiners). Specifically, each observation is a teacher in a given year. The dependent variable is the estimated math value-added (VA) of the teachers in the sample (columns (1) and (5)), the predicted math VA of the teachers (columns (2) and (6)), teachers' experience (columns (3) and (7)), or teachers' education (columns (4) and (8)). The sample is leavers from the school at the end of the year (columns (1)-(4)) and the joiners hired to start the next year (columns (5)-(8)); panel A shows the results using data for both years of the accountability system while panel B displays the results using only the first year. Regressions use a bandwidth of 5 grade points. Standard errors are reported in brackets and clustered at the school level. All regressions control for the "RD running variable" (accountability score relative to the threshold) separately by (year)X(schooltype)X(received lower grade), as well as controls for (year)X(schooltype)X(received lower grade) and all of the main effects and lower-order interactions (e.g., year, schooltype, yearXschooltype, etc.), as well as school controls (which include controls for the average previous year's achievement; previous year's accountability score (second year only); the percent of students that are black, Hispanic, Asian, that receive free and reduced price lunch, and that are immigrants; fixed effects for school size; whether the school was in the NYSBP; and five-year average school turnover prior to the institution of accountability). Data come from the 2007-08 and 2008-09 school years, with the report card grades used being the report card that was received by the school during fall of the school year. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix Table A5. Effect of School Accountability Grades on School Characteristics

Dependent Variable:	Leavers (as % current number teachers)	New hires i.e., joiners (as % current number teachers)	% change in number teachers	% change in enrollment	% change in pupil teacher ratio	Principal turnover
*Independent Var. = School received **lower** grade at the:*	(1)	(2)	(3)	(4)	(5)	(6)
Bottom of the grade distribution						
D/F or C/D thresholds (grouped)	-0.03	-0.01	0.02	0.01	-0.01	0.00
	[0.015]**	[0.015]	[0.018]	[0.029]	[0.032]	[0.079]
N	309	309	309	306	306	290
Dep. Var. Mean	0.14	0.10	-0.03	-0.06	-0.01	0.08
Top of the grade distribution						
B/C or A/B Thresholds (grouped)	0.00	-0.01	-0.01	-0.02	-0.01	0.00
	[0.007]	[0.008]	[0.008]	[0.013]	[0.015]	[0.035]
N	932	932	932	931	931	877
Dep. Var. Mean	0.12	0.09	-0.02	-0.03	-0.01	0.08

Notes. The table presents regression discontinuity estimates of the effect of school accountability grades on school size. Regressions use a bandwidth of 5 grade points. Standard errors are reported in brackets and clustered at the school level. Each observation is a school in a given year. All regressions control for the "RD running variable" (accountability score relative to the threshold) separately by (year)X(schooltype)X(received lower grade), as well as controls for (year)X(schooltype)X(received lower grade) and all of the main effects and lower-order interactions (e.g., year, schooltype, yearXschooltype, etc.) and school controls (which include controls for the average previous year's achievement; previous year's accountability score (second year only); the percent of students that are black, Hispanic, Asian, that receive free and reduced price lunch, and that are immigrants; whether the school was in the NYSBP; fixed effects for school size; and five-year average school turnover prior to the institution of accountability). Columns (1) through (5) estimated weighted by the number of teachers in a school. Data come from the 2008-09 and 2009-10 school years in the New York City Department of Education, using the report card grade that was received by the school during fall of the school year. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix Table A6. Other Robustness Checks for the Regression Discontinuity Turnover Estimates

	Dependent Variable = Teacher Left School						
Independent Var. =	Base sample	Sample excludes phase-out schools	Bounding - includes closing school, assuming 99th percentile turnover	Bounding - includes closing school, assuming 1st percentile turnover	Transfers include mid-year departures	2008 Only	2009 Only
*School received **lower** grade at the:*	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Bottom of the grade distribution							
D/F or C/D thresholds	-0.029	-0.028	-0.029	-0.029	-0.030	-0.027	-0.022
	[0.013]**	[0.013]**	[0.013]**	[0.013]**	[0.014]**	[0.014]**	[0.029]
N	16,897	14,348	16,936	16,936	16,687	11,436	5,461
Top of the grade distribution							
B/C or A/B Thresholds	0.005	0.002	0.005	0.005	0.004	0.010	-0.002
	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.009]	[0.010]
N	54,672	47,919	54,672	54,672	54,131	28,575	26,097

Notes. The table presents regression discontinuity estimates of the effect of school accountability grades on turnover. Regressions use a bandwidth of 5 grade points. Standard errors are reported in brackets and clustered at the school level. Each observation is a teacher in a given year. Columns (1) and (5)-(7) use the base analysis sample which excludes schools undergoing restructuring; column (2) excludes schools that were in the process of being phased out. Columns (3)-(4) perform a bounding exercise to show whether the one closing school affects the results by including the closing school and assuming either turnover at the 99th percentile of the cross-school distribution (42%) or the 1st percentile (0%). The dependent variable is an indicator that the teacher left the school, either between May of one year and November of the following year (all columns except (5)) or between November of one year and November of the following year (column (5)). Column (6) only includes schools from the 2007-08 school year, and column (7) only includes schools from the 2008-09 school year. All regressions control for the "RD running variable" (accountability score relative to the threshold) separately by (year)X(schooltype)X(received lower grade), as well as controls for (year)X(schooltype)X(received lower grade) and all of the main effects and lower-order interactions (e.g., year, schooltype, yearXschooltype, etc.), school covariates (which include controls for the average previous year's achievement; the previous year's accountability score (second year only); the percent of students that are black, Hispanic, Asian, that receive free and reduced price lunch, and that are immigrants; whether the school was in the NYSBP; fixed effects for school size; and five-year average school turnover prior to the institution of accountability), and teacher covariates (which include fixed effects for teacher experience and age, teacher education level, teacher race, and teacher gender). Data come from the 2007-08 and 2008-09 school years in the New York City Department of Education, using the report card grade that was received by the school during fall of the school year. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix Table A7. Main results estimated with vs. without schools directly on threshold

Description	Turnover		Leaver Value-Added				Joiner Value-added			
Dependent Variable	Teacher Left School		Estimated math VA		Predicted math VA		Estimated math VA		Predicted math VA	
Sample:	Full	Non-restructuring	Full	Non-restructuring	Full	Non-restructuring	Full	Non-restructuring	Full	Non-restructuring
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Bottom of the grade distribution (C/D and										
Base sample	-0.029	-0.030	-0.028	-0.150	0.124	0.021	0.869	0.887	0.164	0.214
	[0.013]**	[0.014]**	[0.21]	[0.24]	[0.10]	[0.12]	[0.34]**	[0.38]**	[0.11]	[0.11]**
N	16,897	14,617	479	385	2,228	1,765	126	113	1,493	1,339
Dep. Var. Mean	0.13	0.12	-0.12	-0.09	-0.07	-0.07	-0.12	-0.12	-0.05	-0.04
Sample Excl. Schools on Threshold	-0.029	-0.031	-0.022	-0.145	0.113	0.004	0.904	0.932	0.181	0.234
	[0.014]**	[0.015]**	[0.21]	[0.25]	[0.10]	[0.12]	[0.39]**	[0.42]**	[0.11]*	[0.11]**
N	16,806	14,526	475	381	2,221	1,758	122	109	1,484	1,330
Dep. Var. Mean	0.13	0.12	-0.11	-0.09	-0.07	-0.07	-0.11	-0.12	-0.05	-0.04
Top of the grade distribution										
Base sample	0.005	0.003	-0.130	-0.145	0.119	0.117	-0.392	-0.374	0.023	0.009
	[0.007]	[0.007]	[0.10]	[0.10]	[0.05]**	[0.05]**	[0.20]*	[0.22]*	[0.07]	[0.07]
N	54,672	47,967	1,194	999	5,916	5,056	379	339	4,251	3,765
Dep. Var. Mean	0.11	0.11	-0.05	-0.05	-0.03	-0.01	-0.09	-0.11	-0.05	-0.04
Sample Excl. Schools on Threshold	0.005	0.003	-0.128	-0.143	0.121	0.118	-0.370	-0.354	0.032	0.019
	[0.007]	[0.007]	[0.10]	[0.10]	[0.05]**	[0.05]**	[0.20]*	[0.22]	[0.07]	[0.07]
N	54,585	47,880	1,189	994	5,900	5,040	377	337	4,239	3,753
Dep. Var. Mean	0.11	0.11	-0.05	-0.05	-0.03	-0.01	-0.09	-0.11	-0.05	-0.04

Notes. Table presents the main estimates from the paper, and then re-estimates those results excluding any schools that fall directly on a grade threshold to address concerns with potential biases that could arise from including those schools. Specifically, columns (1) and (2) replicate the main turnover results (columns (1) and (2) from Table 2), and columns (3) - (10) replicate the main joiner and leaver value-added results (columns (1) - (8) of Table 3), both including and excluding the schools that fall directly on a grade threshold. See table notes from Tables 2 and 3 for the details of the specifications including control variables included. All regressions use a bandwidth of 5 grade points. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix Table A8. Main RD results shown separately for the individual thresholds

Description	Turnover		Leaver Value-Added				Joiner Value-added			
Dependent Variable	Teacher Left School		Estimated math VA		Predicted math VA		Estimated math VA		Predicted math VA	
Sample:	Full	Non-restructuring	Full	Non-restructuring	Full	Non-restructuring	Full	Non-restructuring	Full	Non-restructuring
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Bottom of the grade distribution										
C/D and D/F Pooled	-0.029	-0.030	-0.028	-0.150	0.124	0.021	0.869	0.887	0.164	0.214
	[0.013]**	[0.014]**	[0.21]	[0.24]	[0.10]	[0.12]	[0.34]**	[0.38]**	[0.11]	[0.11]**
N	16,897	14,617	479	385	2,228	1,765	126	113	1,493	1,339
Dep. Var. Mean	0.13	0.12	-0.12	-0.09	-0.07	-0.07	-0.12	-0.12	-0.05	-0.04
D/F Threshold	-0.023	-0.049	-0.252	-0.668	0.318	0.353	2.125	3.382	0.096	0.107
	[0.034]	[0.026]*	[0.33]	[0.39]*	[0.16]**	[0.17]**	[0.92]**	[0.53]***	[0.19]	[0.20]
N	5,194	4,325	185	137	876	630	37	34	470	429
Dep. Var. Mean	0.17	0.15	-0.17	-0.11	-0.05	-0.07	-0.17	-0.13	-0.04	-0.05
C/D Threshold	-0.021	-0.022	0.000	-0.068	0.147	0.026	0.251	0.261	0.162	0.194
	[0.014]	[0.015]	[0.22]	[0.24]	[0.10]	[0.11]	[0.43]	[0.44]	[0.12]	[0.12]
N	13,706	12,100	357	306	1,614	1,359	105	95	1,203	1,080
Dep. Var. Mean	0.12	0.11	-0.08	-0.08	-0.10	-0.07	-0.12	-0.14	-0.05	-0.03
Top of the grade distribution										
A/B and B/C Pooled	0.005	0.003	-0.130	-0.145	0.119	0.117	-0.392	-0.374	0.023	0.009
	[0.007]	[0.007]	[0.10]	[0.10]	[0.05]**	[0.05]**	[0.20]*	[0.22]*	[0.07]	[0.07]
N	54,672	47,967	1,194	999	5,916	5,056	379	339	4,251	3,765
Dep. Var. Mean	0.11	0.11	-0.05	-0.05	-0.03	-0.01	-0.09	-0.11	-0.05	-0.04
B/C Threshold	0.003	-0.001	-0.012	-0.035	0.148	0.150	-0.253	-0.333	-0.021	-0.057
	[0.009]	[0.009]	[0.13]	[0.13]	[0.07]**	[0.07]**	[0.29]	[0.35]	[0.09]	[0.09]
N	29,085	24,411	675	544	3,228	2,676	188	162	2,300	1,993
Dep. Var. Mean	0.11	0.11	-0.11	-0.12	-0.06	-0.03	-0.01	-0.04	-0.07	-0.07
A/B Threshold	0.008	0.004	-0.286	-0.277	0.061	0.058	-0.599	-0.470	0.052	0.073
	[0.010]	[0.010]	[0.14]**	[0.15]*	[0.08]	[0.08]	[0.30]**	[0.28]*	[0.10]	[0.10]
N	25,587	23,556	519	455	2,688	2,380	191	177	1,951	1,772
Dep. Var. Mean	0.11	0.10	0.03	0.02	0.00	0.01	-0.16	-0.17	-0.03	-0.01

Notes. Table presents the main estimates from the paper separately for the individual grade thresholds. Specifically, columns (1) and (2) replicate the main turnover results (columns (1) and (2) from Table 2) separately by grade threshold, and columns (3) - (10) replicate the main joiner and leaver value-added results (columns (1) - (8) of Table 3). See table notes from Tables 2 and 3 for the details of the specifications including control variables included. All regressions use a bandwidth of 5 grade points. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix Table A9. Other Robustness Checks for the Regression Discontinuity Joiner Quality Estimates

		Dependent Variable = Math Value-Added				
	Sample:	Base sample	Sample excludes phase-out schools	2008 Only	2009 Only	Sample excludes joiners from failed schools
*Independent Variable = School received **lower** grade at the:*		(1)	(2)	(3)	(4)	(5)
Bottom of the grade distribution						
C/D and D/F thresholds (grouped)		0.87	0.88	0.58	1.14	1.01
		[0.34]**	[0.38]**	[0.43]	[0.68]*	[0.37]***
N		126	112	84	42	122
Top of the grade distribution						
A/B and C/D thresholds (grouped)		-0.39	-0.37	-0.25	-0.18	-0.39
		[0.20]*	[0.22]*	[0.32]	[0.25]	[0.20]*
N		379	339	178	201	375
School and teacher covariates		Yes	Yes	Yes	No	Yes

Notes. Table presents regression discontinuity estimates of the effect of school accountability grades on the quality of joiners hired in the subsequent year, using different samples. Each observation is a teacher in a given year, the dependent variable is the math value-added of the teachers in the sample. Regressions use a bandwidth of 5 grade points. Standard errors are reported in brackets and clustered at the school level. Each observation is a teacher in a given year. Cols (1), (3), (4) and (5) use the base analysis sample; col (2) excludes schools that were in the process of being phased out. Column (3) only includes schools from the 2007-08 school year (so teachers that joined at the beginning of the 2008-09 year), and Column (4) only includes schools from the 2008-09 school year (joiners in 2009-10). Due to a low number of observations, column (4) is estimated without the school and teacher controls, but with all "RD controls" (i.e., (schooltype)X(year)X(received lower grade)). Column (5) excludes joiners who came from failing schools that were closed due to accountability. School controls include controls for the average previous year's achievement; the percent of students that are black, Hispanic, Asian. that receive free and reduced price lunch, and that are immigrants; whether the school was in the NYSBP; fixed effects for school size; and five-year average school turnover prior to the institution

of accountability. Data come from the 2007-08 and 2008-09 school years in the New York City Department of Education, using the report card grade that was received by the school during fall of the school year. $*$ Significant at 10%; $**$ significant at 5%; $***$ significant at 1%.

Appendix Table A10. Descriptive Statistics for the Sample of Schools Not Undergoing Restructuring

	All schools		Not-Restructuring Sample, by Accountability Grade				
	Full sample	Not-Restructuring Sample	A	B	C	D	F
Panel A: Teacher Characteristics							
% Teachers with Master's Degree	0.43	0.44	0.46	0.45	0.43	0.39	0.44
Teacher Experience (years)	9.79	9.81	9.89	10.01	9.69	9.21	9.35
Estimated Math Value-Added	-0.02	-0.01	0.11	0.00	-0.06	-0.11	-0.24
Estimated ELA Value-Added	-0.03	-0.01	0.08	0.00	-0.07	-0.03	-0.16
Predicted Math Value-Added	0.12	0.14	0.17	0.15	0.13	0.06	0.07
Predicted ELA Value-Added	0.09	0.09	0.09	0.08	0.09	0.09	0.01
% Teachers that are:							
Female	0.83	0.84	0.85	0.84	0.84	0.81	0.82
Black	0.20	0.19	0.15	0.18	0.21	0.29	0.22
Non-Hispanic White	0.61	0.63	0.67	0.65	0.61	0.53	0.65
Hispanic	0.14	0.13	0.13	0.12	0.13	0.15	0.09
Asian	0.05	0.04	0.05	0.05	0.04	0.03	0.04
Turnover	0.11	0.11	0.10	0.10	0.11	0.14	0.12
Retirement	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Intra-district transfers	0.04	0.03	0.03	0.03	0.04	0.05	0.05
Exited NYCDOE teacher files	0.07	0.07	0.06	0.06	0.07	0.08	0.07
Sample Size: Teacher-Year Observations (Base Sample)							
All	71,677	62,692	11,263	26,392	18,352	5,728	957
With Math Value-Added Data only	19,092	16,572	2,912	7,081	4,784	1,569	226
Sample Size: Unique Teachers (Base Sample)							
All	50,616	44,303					
With Math Value-Added Data only	13,202	11,499					
Panel B: School Characteristics							
Enrollment	809	783	656	685	667	567	492
% Students that are:							
Black	0.32	0.32	0.28	0.32	0.38	0.46	0.39
Non-Hispanic White	0.15	0.17	0.17	0.17	0.16	0.09	0.18
Hispanic	0.40	0.37	0.37	0.38	0.36	0.40	0.37
Asian	0.12	0.13	0.18	0.13	0.09	0.04	0.06
Immigrants	0.02	0.02	0.01	0.02	0.01	0.01	0.01
Components of Accountability Grades							
Environment Score	7.43	7.58	9.33	8.12	6.97	6.05	6.54
Performance Score	15.11	15.39	18.62	15.86	14.12	11.46	9.65
Progress Score	26.67	26.54	34.31	29.33	21.75	16.02	11.63
Additional Credit	2.18	2.13	3.17	2.31	1.17	0.72	0.24
Overall Score	51.40	51.65	65.45	55.64	44.02	34.26	28.07
Other characteristics							
NY School Bonus Program	0.15	0.13	0.13	0.15	0.14	0.17	0.18
Being phased out[a]	0.01	0.01	0.00	0.00	0.00	0.04	0.05
Not restructuring at baseline[b]	0.87	1.00	1.00	1.00	1.00	1.00	1.00
Sample Size: Schools							
Number of school-year observations	1,243	1,123	207	456	323	115	22
Number of unique schools	847	766					

Notes. Data come from the 2007-08 and 2008-09 school years in the New York City Department of Education. The accountability grade is the school report card grade that was received by the school during fall of the school year. Sample here limited to the sample used for the base turnover analysis: schools within a 5-point bandwidth of one of the grade thresholds. For all columns except the first, sample also limited to schools that were not undergoing restructuring prior to the institution of the accountability system (i.e., that were not in year 2+ of restructuring in the 2007-08 school year).

a. Phase outs proxied for by schools that received accountability grades in one year but not the subsequentyear.

b. Schools that were not undergoing restructuring prior to the institution of the accountability system (i.e., that were not in year 2+ of restructuring in the 2007-08 school year).

Appendix Table A11. Correlates of teacher mobility

Dep. Var.:	Teacher left school	
Independent Variables:	(1)	(2)
Years of experience	-0.00208***	-0.00125***
	[0.000254]	[0.000227]
Has masters	-0.0180***	-0.0126***
	[0.00324]	[0.00282]
Female	-0.0410***	-0.00860***
	[0.00443]	[0.00322]
Black	0.00424	-0.0424***
	[0.00502]	[0.00390]
Hispanic	-0.0164***	-0.0406***
	[0.00427]	[0.00420]
School Fixed Effects?	No	Yes

Notes. Table presents regressions of teacher mobility on teacher characteristics. Each observation is a teacher in a given year. Data comes from the pre-accountability era (2007 school year). Standard errors clustered at the school level.

* Significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix 1

Value-Added Estimation

To estimate teacher value-added, I follow an approach that has been experimentally validated in the economics of education literature (Kane and Staiger 2008) and estimate the following regression using the matched student-teacher panel:

$$A_{ijgt} = \alpha + \beta_1 A_{i,j-1,g-1,t-1} + \beta_2 \bar{A}_{-i,j-1,g-1,t-1} + \beta_3 X_i + \tau_t + \tau_g + \tau_j + \eta_{jt} + \varepsilon_{ijgt} \quad (2)$$

where A_{ijgt} is the achievement score (either mathematics or English Language Arts, standardized by year and grade) of student i in the classroom of teacher j in grade g and year t; $A_{i,j-1,g-1,t-1}$ is the student's lagged achievement; $\overline{A}_{-i,j-1,g-1,t-1}$ represents the average previous-year achievement of student i's classmates (to control for peer effects); X_i are student demographics (e.g., gender, ethnicity, eligibility for free-and-reduced-price-lunch); the τ terms represent fixed effects for the year, the grade, and the teacher, respectively; and η_{jt} and ε_{ijgt} represent classroom-level and individual-level error terms, both mean zero and assumed to be independently and identically distributed over time. Following the literature (e.g., Jackson 2012), in order to have estimates that are comparable across schools, I omit school fixed effects, but the results are quantitatively very similar if I instead estimate using school fixed effects. I only use data from years before the institution of accountability in order to isolate teacher quality from teacher responses to accountability.

I then follow the approach outlined in Kane and Staiger (2008) and Jackson (2009) to create Empirical Bayes (EB) estimates of teacher value-added. Although the estimates obtained by

estimating equation 2 directly are consistent (under identifying restrictions), they are not efficient. EB estimates are more efficient, providing the Best Linear Predictor of the random teacher effect in equation 2, which is also the posterior mean with normally distributed errors. Consider the error term in equation 2, $w_{ijgt} \equiv \tau_j + \eta_{jt} + \varepsilon_{ijgt}$. It is the sum of the teacher effect, assumed constant across years, a mean-zero year-specific classroom error, and a mean-zero year-specific student error. To construct EB estimate, I need to estimate the variance of each component. To do this, I first estimate equation 2 using OLS. For the teacher effect, I calculate the mean residual, by teacher, in each year, and use the covariance between these residuals in adjacent years as the estimate of the variance of the teacher effect, $\hat{\sigma}_\tau^2 = Cov(\overline{w}_{jgt}, \overline{w}_{jgt-1})$.[i] For the variance of the student effect, I calculate the variance of the student residuals after the classroom mean residual has been removed: $\hat{\sigma}_\varepsilon^2 = Var(w_{ijgt} - \overline{w}_{jgt})$. Finally, under the assumption that all three components of the error term are orthogonal to each other, I calculate the variance of the classroom term as the variance of the total error term minus the variance of the teacher and student components: $\hat{\sigma}_\eta^2 = Var(w_{ijgt}) - \hat{\sigma}_\tau^2 - \hat{\sigma}_\varepsilon^2$.

Next, I compute a raw estimate of a teacher's effect as a weighted average of their classroom residuals ($\overline{w}_{jgt}$), where each classroom is weighted by the inverse of its variance: $\hat{\tau}_j = \sum_{t=1}^{T_j} \overline{w}_{jgt} \frac{(\sigma_\eta^2 + \sigma_\varepsilon^2/N_{jt})^{-1}}{\sum_{t=1}^{T_j}(\sigma_\eta^2 + \sigma_\varepsilon^2/N_{jt})^{-1}}$, where N_{jt} is the number of students in classroom jt and T_j is the total number of classrooms for teacher j.

Finally, I weight this estimate by an estimate of the precision of the teacher's effect to form the empirical Bayes estimate: $\hat{\tau}_j^{EB} = \hat{\tau}_j \frac{\sigma_\tau^2}{\sigma_\tau^2 + [\sum_{t=1}^{T_j}(\sigma_\eta^2 + \sigma_\varepsilon^2/N_{jt})^{-1}]^{-1}}$.

Since the identification of true teacher value-added depends on strong identification

assumptions, e.g., that assignment of students to teachers is orthogonal to the student error term ε_{ijgt} in equation 2, recent literature has highlighted the potential biases of value-added measures (e.g., Rothstein 2010). However, given the RD framework, my identification requirements are less stringent than if I was, say, trying to evaluate teachers based on the estimates. The RD results would only be biased if, conditional on the accountability score, there were differences in the average school-level bias of the value-added estimates that was correlated with the grades. Since the value-added was calculated using pre-period data, this is unlikely. Of greater concern is the comprehensiveness of the value-added estimates: if there are aspects of teacher quality which are not summarized well in teacher value-added measures (which is likely), then my analysis will not incorporate these aspects.

A.1 Calculation of Predicted VA

To compute teacher predicted VA, I follow the approach used in Jackson (2012) and estimate an equation of student achievement (with the inclusion of observable teacher characteristics) on students, using the same period of data used above for calculating teacher VA. Specifically, I estimate the following equation:

$$A_{ijgt} = \alpha + \beta_1 A_{i,j-1,g-1,t-1} + \beta_2 \bar{A}_{-i,j-1,g-1,t-1} + \beta_3 X_i + \beta_4 W_{jt} + \beta_5 W_j + \tau_t + \tau_g + \eta_{jt} + \varepsilon_{ijgt} \quad (3)$$

where all variables are defined as before, and the equation now omits the teacher fixed effect τ_j but includes W_{jt} which captures time-varying teacher characteristics (i.e., experience), and W_j which is a vector of observable teacher characteristics (e.g., education and demographics). Using the estimates from equation (3), I predict teacher effectiveness using the observable teacher characteristics. Specifically, the predicted VA for teacher j is $\hat{\beta}_4 W_{jt} + \hat{\beta}_5 W_j$ (i.e., the predicted VA associated with teacher j's experience, education, and demographics). This measure serves

as a useful summary statistic for all of the observable teacher characteristics. It is a weighted average of a teacher's observable characteristics, where the weights are determined by the characteristics' relationship with actual student achievement. The advantage of this measure is that it can be calculated for all teachers irrespective of when they enter the data.[ii] Unfortunately, I do not have as extensive observable teacher characteristics as some previous papers (e.g., Jackson 2012) to be able to do the prediction.

Appendix 2

Regression Discontinuity Bandwidth Selection

Since there is no universally agreed-upon method for determining bandwidth for an RD analysis, I follow the standard approach of examining the robustness of the results to different bandwidths.

To select the base bandwidth used for the analyses, I follow the "leave one out" cross-validation procedure of Ludwig and Miller (2007) and Lee and Lemieux (2010) in which I estimate locally linear models at different bandwidths while omitting one observation, calculate the cross-validation criterion as the average squared difference between the predicted and actual values for the omitted observations, and choose the bandwidth that minimizes the cross-validation criterion. The optimal bandwidth using covariates for both the upper and lower thresholds was 5 and so I use that as my base bandwidth for the regressions. When looking at the value-added outcomes and using as my samples either the joiners or the leavers, the optimal bandwidths ranged from 2-7, with the median 4. For consistency, I adopt the turnover bandwidth (5) for my base bandwidth and then show robustness to different bandwidths in the robustness section (Section VI, Tables 9 and 10).

To check robustness, I also calculate a version of the Imbens-Kalyanarman (IK) optimal bandwidth (Imbens and Kalyanaraman 2012).[iii] For the turnover outcomes, these range from 4-6 with a median of 5, and for the value-added, they range from 2-5 with a median of 3; all of these are in the ranges of bandwidths displayed in the robustness tables (Tables 9 and 10).

For the graphs, I show a bandwidth two times wider than the base bandwidth used in the regressions to give a better sense of the regression function.

---

[i] This is slightly different from the procedure used by Kane and Staiger (2008) and Jackson (2009), who use the covariance between adjacent classroom-level residuals instead of teacher-level residuals since they both use elementary data only in which the majority of teachers only teach one classroom.

[ii] I have test score data from both elementary and middle schools so the measure is defined for all school levels.

[iii] The IK formula is not developed for the pooled threshold model I use here, where I interact the running variable for indicators for which threshold (school type and year) a given observation is at. I try two modifications to the IK procedure to try to get reasonable estimates within the pooled setting. First, I simply ignore the fact that I am pooling across thresholds, so calculate the IK bandwidth that would be appropriate if there were no interactions with the running variable. Second, I calculate the IK bandwidth separately for each threshold (i.e., for each school type and year). I then calculate the weighted average of the separate bandwidths (weighted by the relative sample sizes), where all are normalized by their sample sizes. Finally, I normalize the averaged bandwidth by the total sample size. In practice, the two methods yield nearly identical results.