Appendix

For online publication only

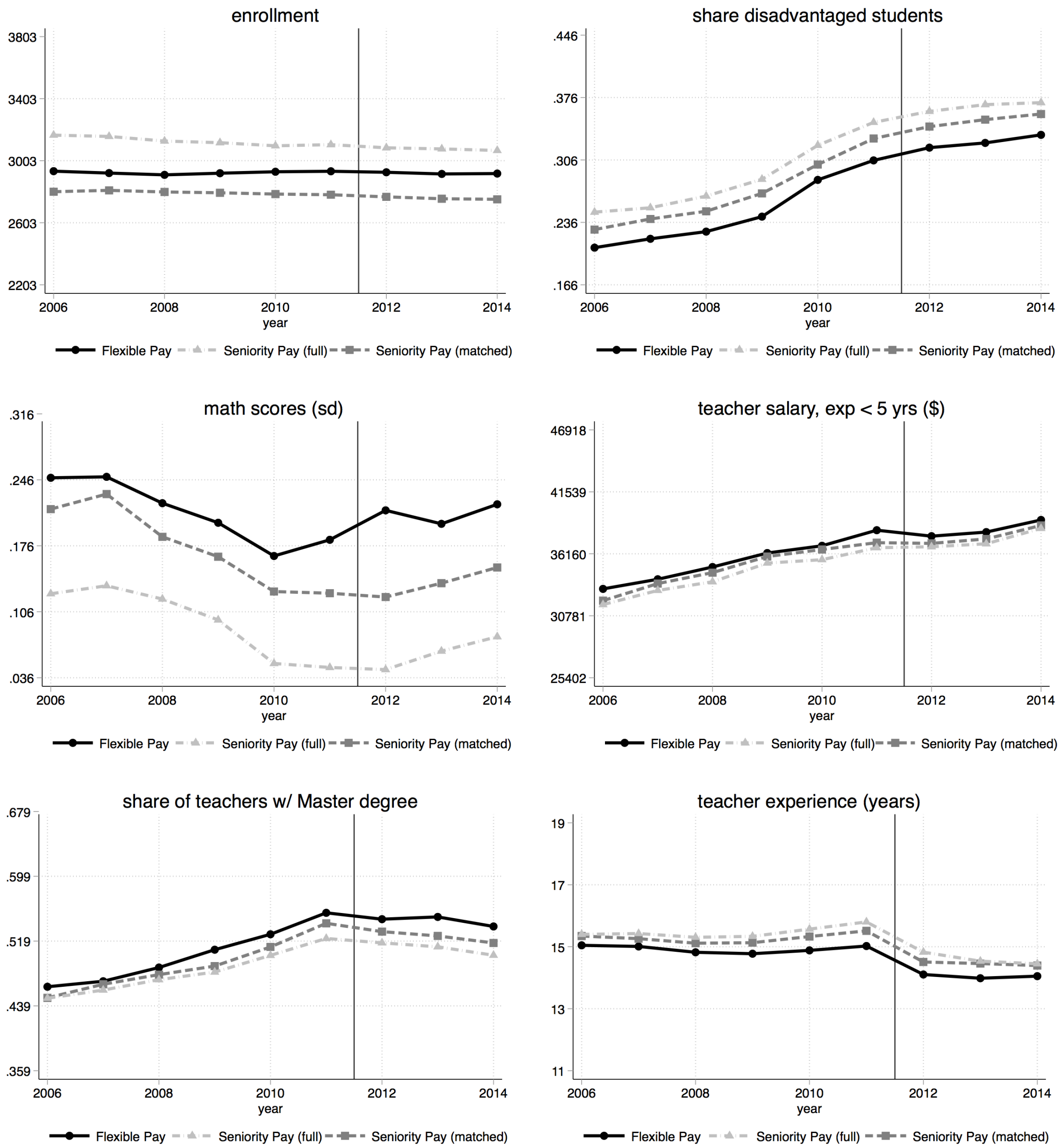
Appendix A Additional Tables and Figures

Figure A1: Salary Schedule - Racine School District, 2016

Step	BA	BA+12	BA+24	MA
1	40,593	42,784	44,976	47,169
2	41,526	43,717	45,909	48,516
3	42,459	44,651	46,842	49,864
4	43,392	45,584	47,775	51,211
5	44,325	46,517	48,709	52,560

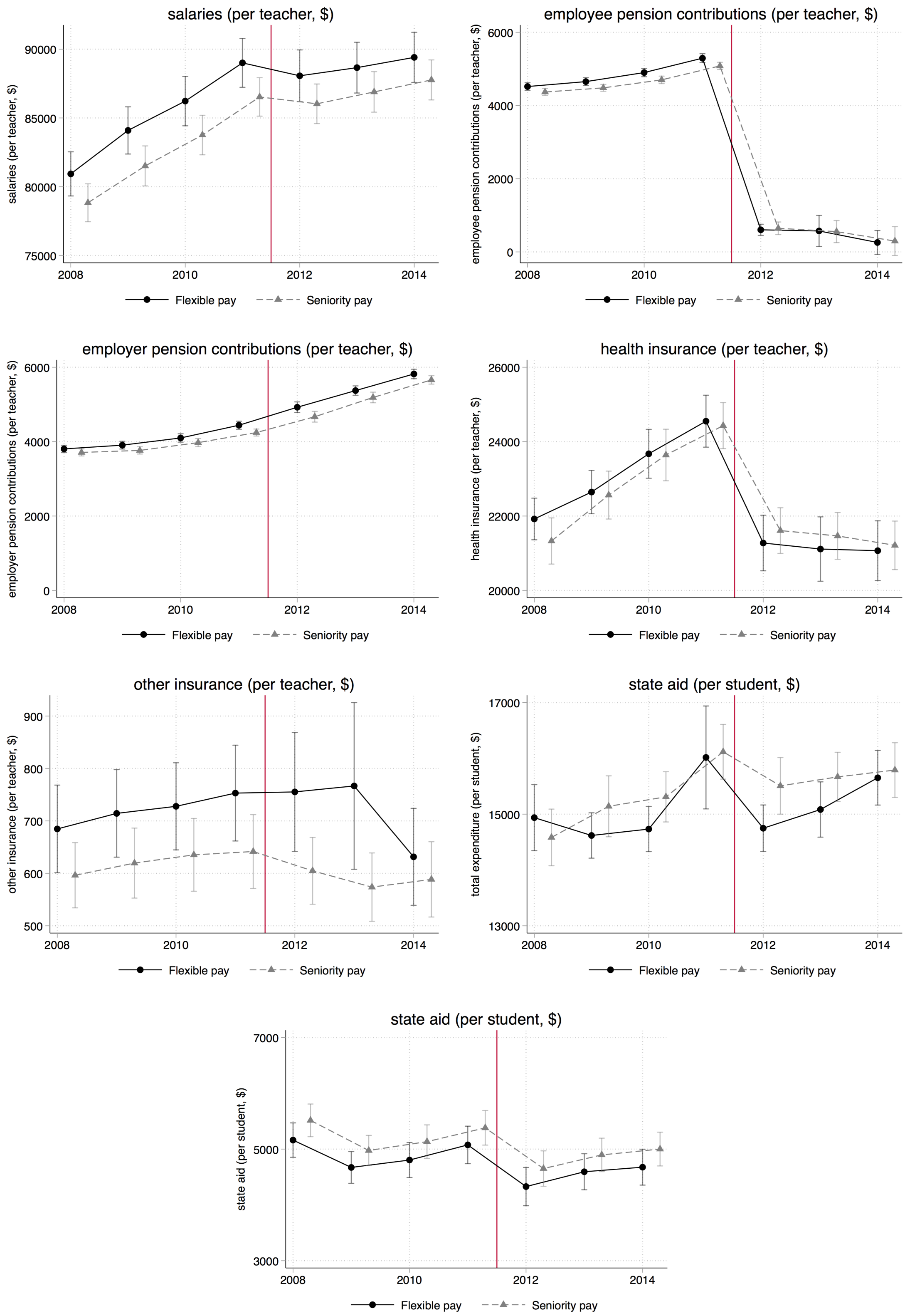
Notes: Subsection of the salary schedule Used by the school district of Racine in 2011. Source: http://www.rusd.org.

Figure A2: Trends in Observable Characteristics of School Districts, 2006–2014



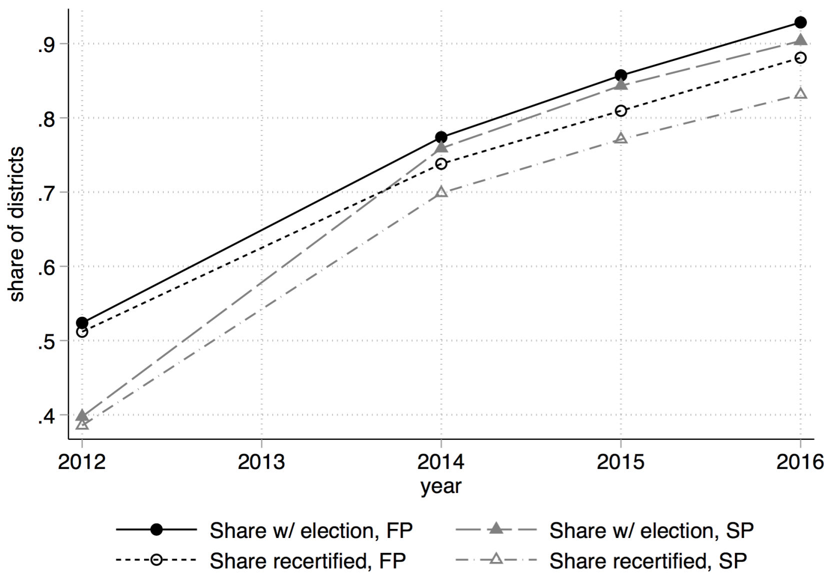
Notes: Trends in average characteristics of school districts over time, separately for FP, SP, and matched SP districts. From top left to bottom right: enrollment, share of economically disadvantaged students, math test scores, salary for teachers with less than 5 years of experience, share of teachers with a Master degree, teacher experience. Standard errors in parentheses are clustered at the district level. The matched sample is obtained using nearest-neighbor matching on observable characteristics of the school districts.

Figure A3: Components of Districts' Budgets, 2008–2014



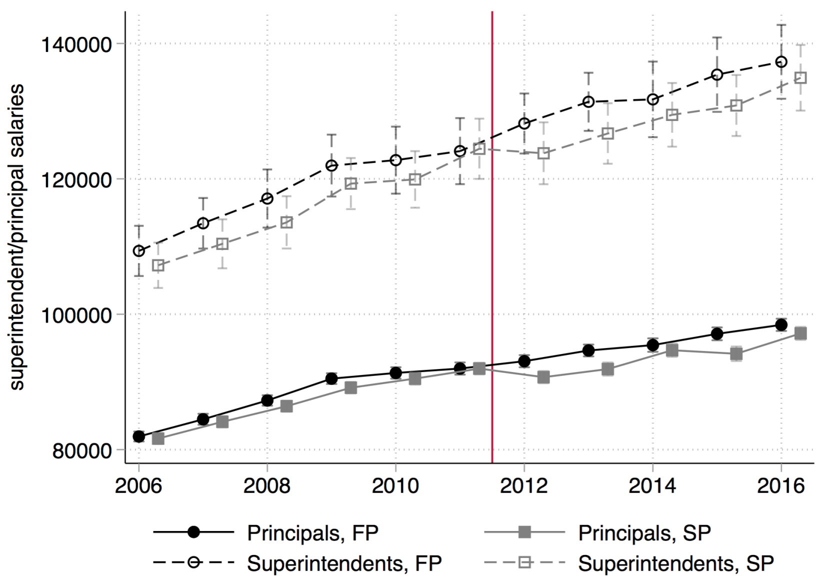
Notes: Means and 90% confidence intervals of different types of district expenditures and revenues, over time and by type of district. From top-left to bottom-right: expenditure on salaries (per teacher), employee and employer contributions to the pension fund (per teacher), expenditure on health care plans (per teacher), expenditure on other type of insurance (per teacher), total expenditure (per student), and state aid (per student).

Figure A4: Districts with Union Recertification Elections, by Year and Type of District: Overall Share and Share Recertified, 2012–2016



Notes: Share of districts which had a union recertification election in each year, and share of districts where the election was successful and the union re-certified, separately for FP and SP districts. In 2013 no elections were held due to ongoing litigation over the recertification requirements of Act 10.

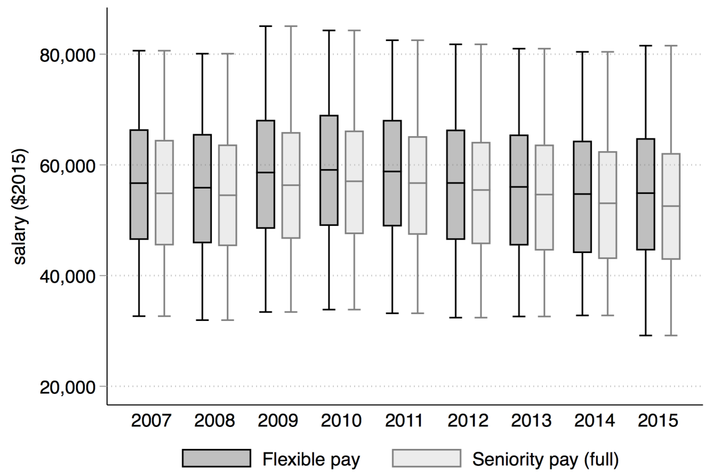
Figure A5: Salaries of District Superintendents, 2012–2016



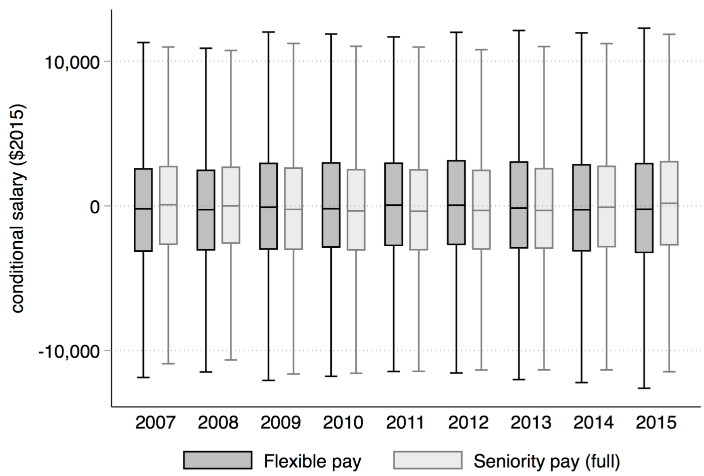
Notes: Means and 90% confidence intervals of salaries of district superintendents and school principals, by type of district (FP vs. SP) and over time.

Figure A6: Salary Distribution in FP and SP districts

Panel A: Raw salaries

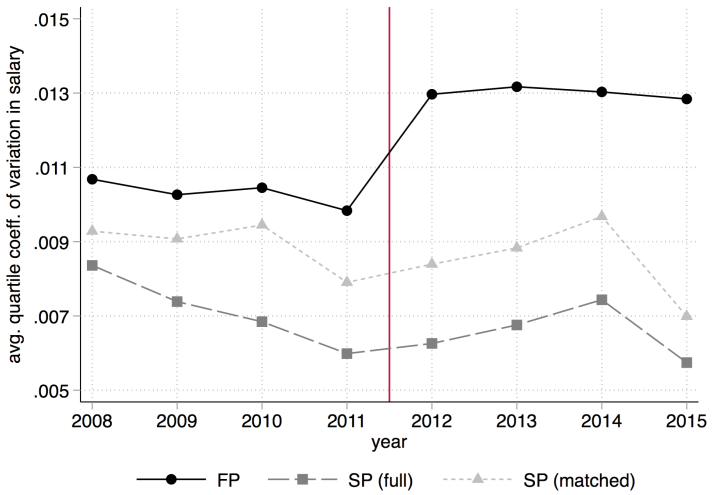


Panel B: Conditional salaries



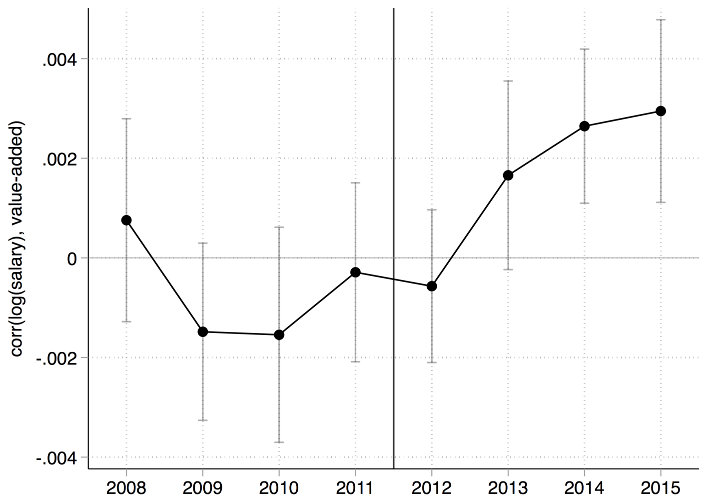
Notes: Box plots of salaries (Panel A) and conditional salaries (Panel B) in FP and SP districts for the years 2007–2015. Conditional salaries are the residuals of a regression of salaries on a non-parametric function of years of experience, interacted with indicators for the highest education degree and with a dummy for years after 2011; district fixed effects interacted with a dummy for years after 2011; and year fixed effects. The sample is restricted to tenured teachers (with more than 3 years of experience) working in FP and SP districts.

Figure A7: Quartile Coefficient of Dispersion in Salaries, 2008–2015



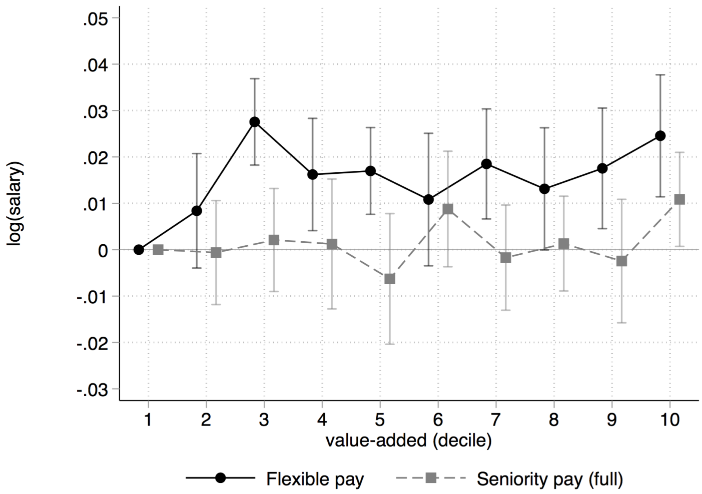
Notes: Trends in the median district-level quartile coefficient of dispersion in salaries. Quartile coefficients of dispersion are calculated as the ratio between the difference and the sum of the 75th and 25th percentiles of salaries, computed separately for each group of teachers with the same experience (in 2-years bins) and highest education degree in each district. The sample is restricted to tenured teachers (with more than 3 years of experience) working in FP, SP, and matched SP districts. The matched sample of SP districts is obtained via nearest-neighbor matching on observable characteristics of the school districts.

Figure A8: Correlation, Salaries and Value-Added: All Wisconsin districts, 2008–2015



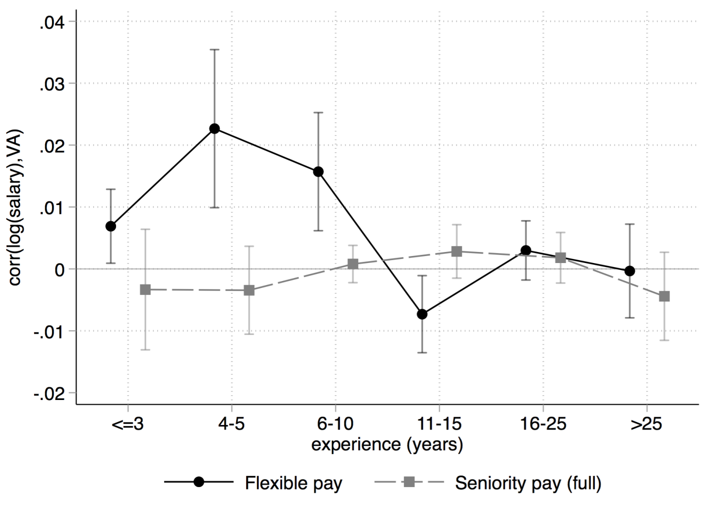
Notes: OLS estimates and 90% confidence intervals of the coefficients δ_s in the regression $log(w_{ijt}) = \sum_{s=2008}^{2015} \delta_s \tau_s * VA_{it} + \beta X_{it}^w + \theta_j + \tau_t + \varepsilon_{ijt}$. The variable $log(w_{ijt})$ is the natural logarithm of salary for teacher i working in district j in year t. The variable VA_{it} is teacher VA. The vector X_{it}^w includes a non-parametric function of years of experience, interacted with indicators for the highest education degree and with a dummy for years after 2011. The vector θ_j contains district fixed effects, and the vector τ_t contains year fixed effects. The coefficients δ_s are estimated separately for FP and SP districts. VA is calculated as the average of a time-varying measure over the years 2007–2011 and 2012–2015. The sample is restricted to tenured teachers (with more than 3 years of experience) working in all Wisconsin districts. Bootstrapped standard errors are clustered at the district level.

Figure A9: Salaries, by Decile of Value-Added



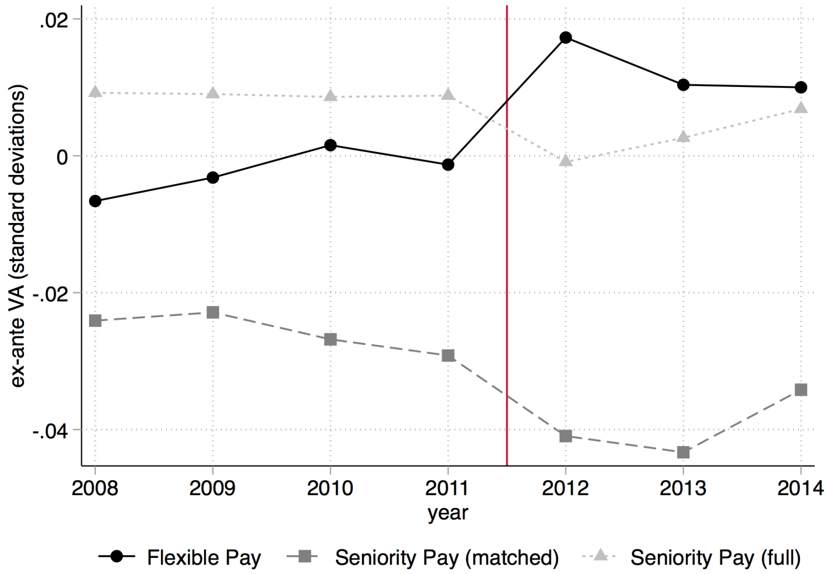
Notes: OLS estimates and 90% confidence intervals of the coefficients δ_s in the regression $log(w_{ijt}) = \sum_{s=1}^{10} \delta_s^0 \mathbb{1}(D(VA_{it}) = s) + \sum_{s=1}^{10} \delta_s \mathbb{1}(D(VA_{it}) = s) * \mathbb{1}(t > 2011) + \beta X_{it}^w + \theta_j + \tau_t + \varepsilon_{ijt}$. The variable $log(w_{ijt})$ is the natural logarithm of salary for teacher i working in district j in year t. The variable VA_{it} is teacher VA. The function $D(VA_{it})$ denotes the decile in the distribution of value added, and $\mathbb{1}(.)$ is an indicator function. The vector X_{it}^w includes a non-parametric function of years of experience, interacted with indicators for the highest education degree and with a dummy for years after 2011. The vector θ_j contains district fixed effects, and the vector τ_t contains year fixed effects. The coefficients δ_s are estimated separately for FP and SP districts. VA is calculated as the average of a time-varying measure over the years 2007–2011 and 2012–2015. The sample is restricted to tenured teachers (with more than 3 years of experience) working in FP and SP districts. Bootstrapped standard errors are clustered at the district level.

Figure A10: Correlation, Salaries and Value-Added: by Experience



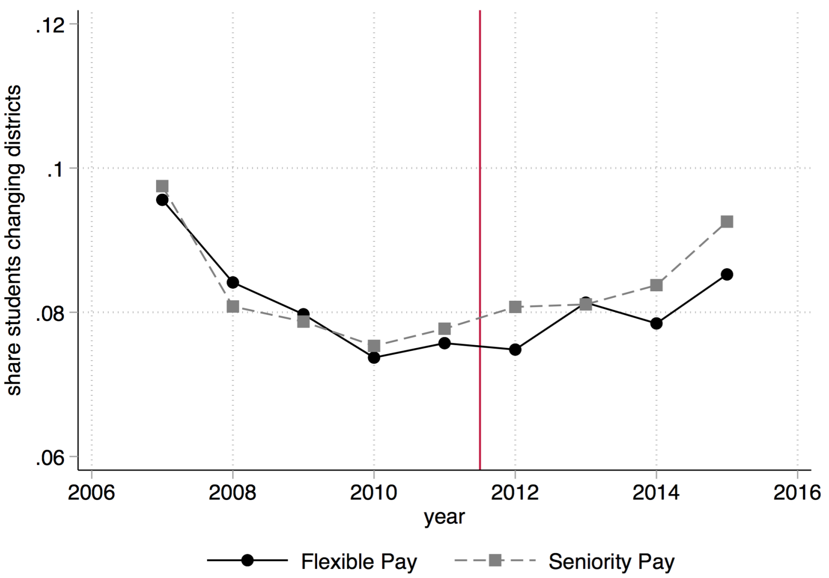
Notes: OLS estimates and 90% confidence intervals of the coefficients δ_s in the regression $log(w_{ijt}) = \sum_{s=1}^{6} \delta_s^0 \mathbb{1}(exp_{it} = s)q_{it} + \sum_{s=1}^{6} \delta_s \mathbb{1}(exp_{it} = s)VA_{it}\mathbb{1}(t > 2011) + \tau_t + \beta X_{it}^w + \varepsilon_{ijt}$. The variable $log(w_{ijt})$ is the natural logarithm of salary for teacher i working in district j in year t. The variable VA_{it} is teacher VA, the variable exp_{it} is a categorical function of years of experience, where the categories are ≤ 3, 4-5,6-10,11-15, 16-20, and >20 and $\mathbb{1}(.)$ is an indicator function. The vector X_{it}^w includes a non-parametric function of years of experience, interacted with indicators for the highest education degree and with a dummy for years after 2011. The vector θ_j contains district fixed effects, and the vector τ_t contains year fixed effects. The coefficients δ_s are estimated separately for FP and SP districts. VA is calculated as the average of a time-varying measure over the years 2007–2011 and 2012–2015. The sample is restricted to tenured teachers (with more than 3 years of experience) working in FP and SP districts. Bootstrapped standard errors are clustered at the district level.

Figure A11: Changes in the Composition of the Teaching Workforce: Ex Ante Value-Added, FP vs. SP, 2008–2015



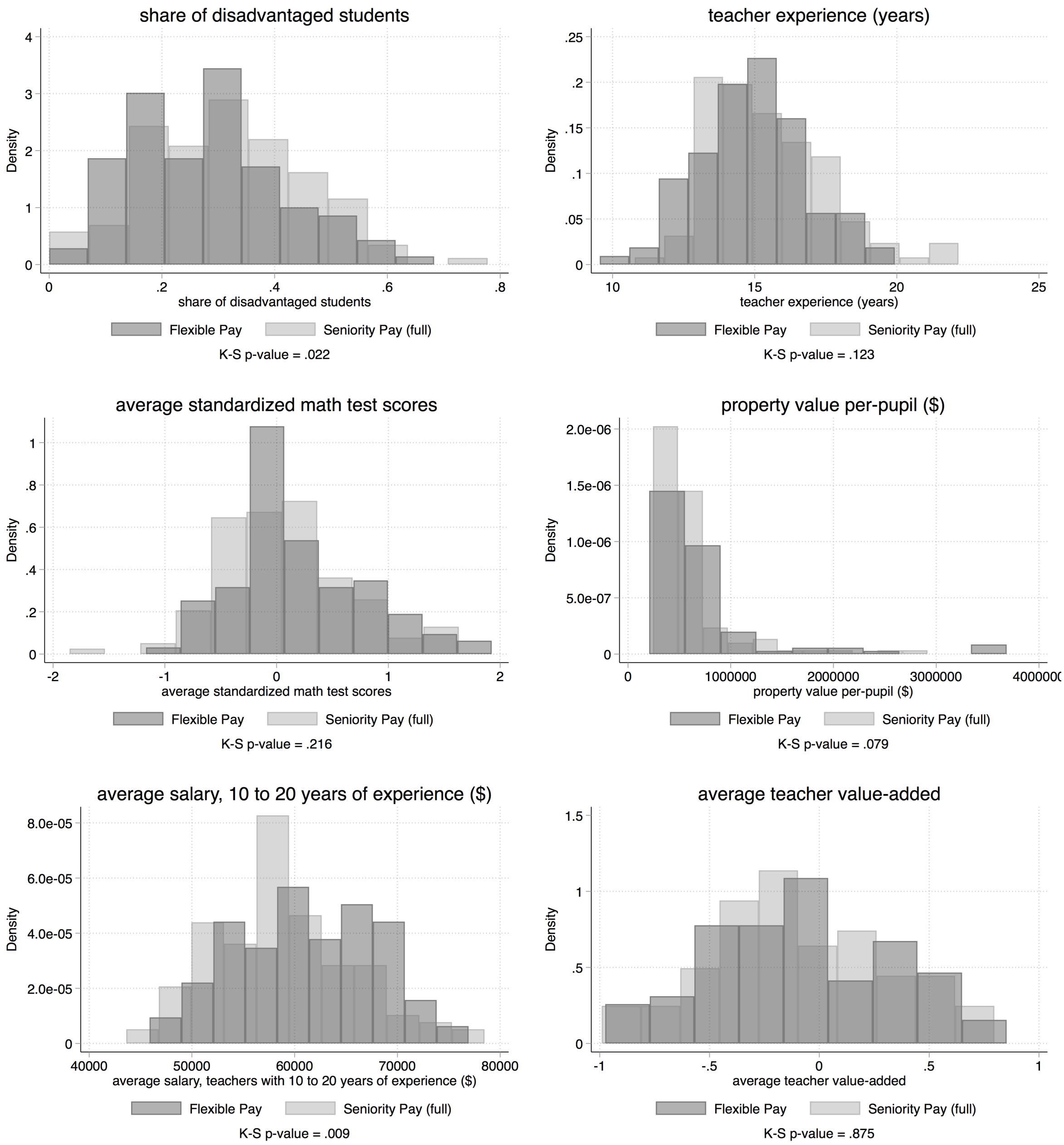
Notes: Average *ex ante* VA of teachers working in FP, SP, and matched SP districts between 2008 and 2015. *Ex ante* VA is calculated as the average of a time-varying measure over the years 2007–2011. The sample is restricted to tenured teachers (with more than 3 years of experience). The matched sample is obtained using nearest-neighbor matching on observable characteristics of the school districts.

Figure A12: Student Movements Across Districts



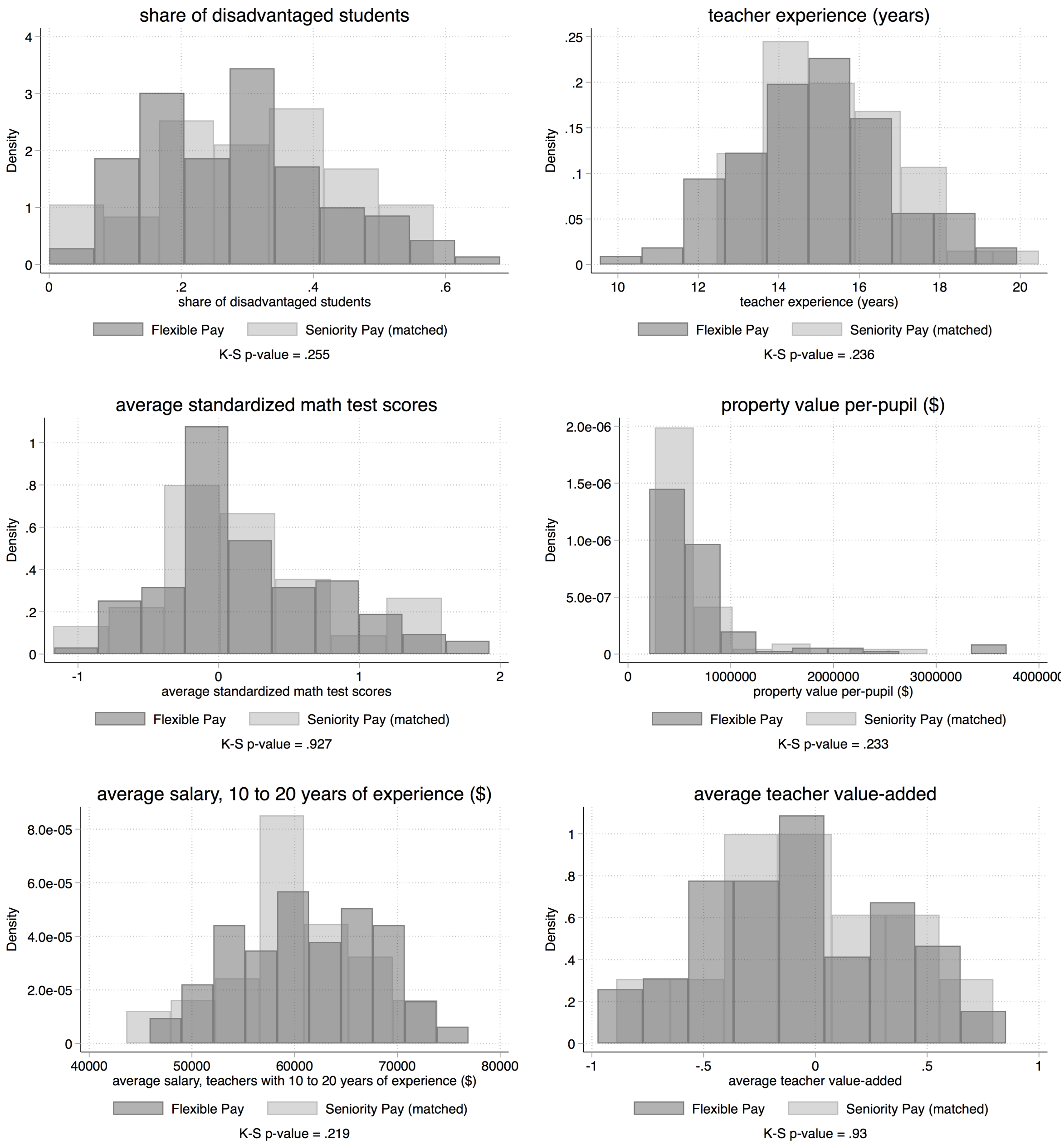
Notes: Share of students in grades 4-8 who change district, by type of district of destination. Each share is defined over the total number of students in each year and type of district, enrolled in FP and SP districts.

Figure A13: Distribution of Average District-Level Characteristics - Full sample



Notes: Distribution of district-level characteristics of FP and SP districts in the period 2007–2011. P-values of a Kolmogorov-Smirnov test of equality in distributions are reported below each graph. The FP subsample includes 102 districts. The SP subsample includes 122 districts.

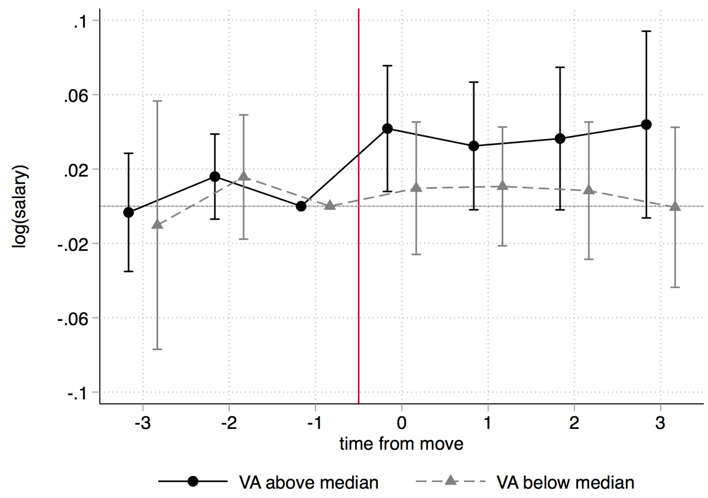
Figure A14: Distribution of Average District-Level Characteristics - Matched sample



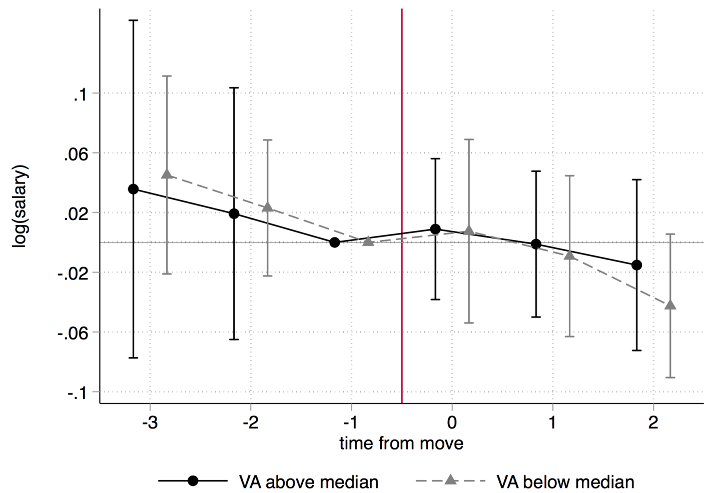
Notes: Distribution of district-level characteristics of FP and matched SP districts in the period 2007–2011. P-values of a Kolmogorov-Smirnov test of equality in distributions are reported below each graph. The FP subsample includes 102 districts. The matched sample is obtained using nearest-neighbor matching on observable characteristics of the school districts, and includes 56 districts.

Figure A15: Salaries of Movers Before and After A Move - Matched Sample

Panel A: Flexible pay

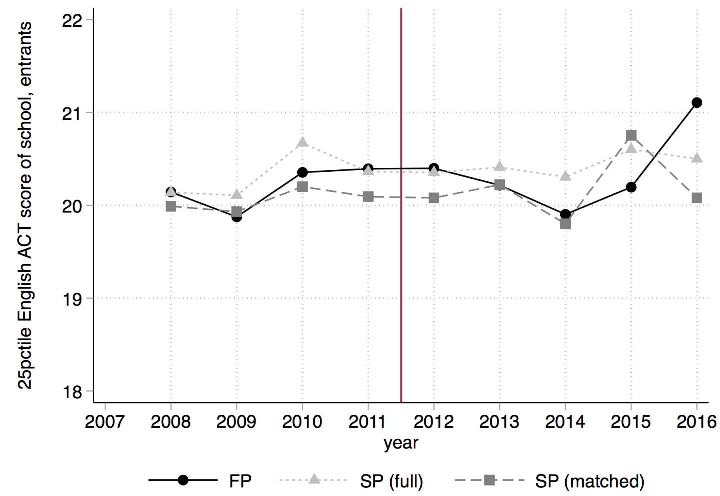


Panel B: Seniority pay



Notes: OLS estimates and 90% confidence intervals of the coefficients β_s in the regression $\log(w_{ijt}) = \sum_{k=-3}^{3} \beta_k^0 \mathbb{1}(t - Y_i^m = k) + \sum_{k=-3}^{3} \beta_k \mathbb{1}(t - Y_i^m = k) * \mathbb{1}(Y_i^m > 2011) + \gamma X_{it}^w + \theta_j + \tau_t + \varepsilon_{ijt}$. The variable $log(w_{ijt})$ is the natural logarithm of salaries for teacher i working in district j in year t. The variable Y_i^m denotes the year in which teacher i moves to district j,$\mathbb{1}(.)$ is an indicator function, and the vector X_{it}^w includes a non-parametric function of years of experience, interacted with indicators for the highest education degree and with a dummy for years after 2011. θ_j are district fixed effects and τ_t are year fixed effects. The coefficient β_{-1} is normalized to 0. The parameters are estimated separately for teachers in FP and in matched SP districts, with *ex ante* VA above and below the median. The sample is restricted to tenured teachers (with more than 3 years of experience) working in FP and matched SP districts. The matched sample of SP districts is obtained via nearest-neighbor matching on observable characteristics of the school districts. Standard errors are clustered at the district level.

Figure A16: New Teachers: Share (top panel) and Selectivity of College Institution (bottom panel)



Notes: Top panel: Share of new teachers, by type of district. Each share is defined over the total number of Wisconsin teachers in each year and type of district. Bottom panel: 25-percentile English ACT score of the institution where a teacher obtained her latest degree. Institutions only reporting SAT scores are assigned the corresponding ACT score based on the score distribution. The sample is restricted to teachers with college degrees from institutions requiring ACT or SAT scores.

Table A1: Difference-In-Difference, Characteristics of Wisconsin School Districts:Fp vs. SP, 2008–2015

	Students				Superintendents	
	Nr	Black	low SES	Math score	Salary	Experience
FP	-188.436	-0.003	-0.041**	0.114	2208.871	-1.079
	(718.895)	(0.008)	(0.018)	(0.078)	(2933.341)	(1.182)
FP * post	32.450	0.001	0.001	0.035	1892.462	-0.543
	(52.769)	(0.001)	(0.004)	(0.027)	(1469.699)	(1.127)

	Principals		Teachers			
	Salary	Experience	Salary	Experience	Masters	Value-added
FP	2460.725**	0.318	2107.311***	-0.627**	0.025	0.047
	(1107.724)	(0.688)	(696.609)	(0.263)	(0.019)	(0.063)
FP * post	829.237	-0.592	23.405	0.070	0.009	0.007
	(628.673)	(0.633)	(285.354)	(0.200)	(0.009)	(0.019)

	Districts					
	Urban	Suburban	Expend. pp	Aid pp	ln(property)	N Admins/T
FP	-0.005	0.097*	-180.527	-312.253	0.157**	-0.003
	(0.035)	(0.058)	(338.640)	(252.811)	(0.072)	(0.009)
FP * post			-193.691	13.392	0.004	-0.004
			(323.156)	(69.418)	(0.009)	(0.005)

	Budget (Expenditure per Pupil)				Unions	
	Salaries	Retirement	Health ins.	Other ins.	Held election	Recertified
FP	-12.652	1.870	-36.442	4.000	-0.000	-0.000
	(94.860)	(5.943)	(46.783)	(4.085)	(0.000)	(0.000)
FP * post	16.071	2.862	-10.077	3.647	0.047	0.004
	(59.233)	(6.221)	(34.926)	(3.184)	(0.034)	(0.018)

Notes: The table shows estimates of α and β in $X_{dt} = \alpha FP_d + \beta FP_d * post_t + \gamma post_t + \varepsilon_{dt}$, where X_{dt} includes a range of district attributes, including (in order from top-left to bottom-right): the number of students, the share of students who are Black and low-SES, math test scores; district superintendents' salary and years of experience; principals' salary and years of experience; teachers' salary, years of experience, share with a Master's, and VA; indicators for whether the district is located in an urban or suburban area; expenditure, state aid, and the logarithm of property values per pupil; the total number of administrative staff per teacher; per pupil expenditure on salaries, retirement, health insurance, and other insurance; and indicators for whether a district held a union recertification election and for whether the election was won. The variable FP_d equals 1 for FP districts, and $post_t$ equals 1 for years after Act 10. Each observation corresponds to a school district in a given year. The FP subsample includes 102 districts, the SP subsample includes 122 districts. The subsample covers 83 percent of the total student population.

Table A2: Summary Statistics, District Management, 2009–2011

	(1) Full sample			(2) Matched sample		
	FP	SP	Difference	FP	SP	Difference
admin/staff per teacher	0.24	0.24	-0.0016	0.24	0.23	0.0085
			(0.0090)			(0.010)
female superintendent	0.21	0.13	0.077*	0.21	0.12	0.088
			(0.045)			(0.058)
superintendent age	54.6	54.1	0.50	54.6	53.3	1.28
			(0.86)			(1.09)
superintendent experience	25.3	26.4	-1.12	25.3	25.8	-0.51
			(1.25)			(1.60)
superintendent holds BA	0.0033	0	0.0033	0.0033	0	0.0033
			(0.0030)			(0.0044)
superintendent holds Master/PhD	0.80	0.80	0.0034	0.80	0.76	0.041
			(0.050)			(0.062)
superintendent salary ($)	123037.7	121118.5	1919.2	123037.7	124798.6	-1760.8
			(2985.8)			(3556.4)
new superintendent	0.99	1	-0.0065	0.99	1	-0.0065
			(0.0042)			(0.0062)
female principal	0.40	0.36	0.039	0.40	0.37	0.037
			(0.034)			(0.039)
principal age	48.4	48.2	0.26	48.4	47.8	0.66
			(0.62)			(0.71)
principal experience	20.8	20.4	0.32	20.8	20.3	0.42
			(0.71)			(0.83)
principal holds BA	0.014	0.021	-0.0070	0.014	0.020	-0.0059
			(0.0090)			(0.0078)
principal holds Master/PhD	0.95	0.94	0.013	0.95	0.93	0.015
			(0.022)			(0.024)
principal salary ($)	89494.1	87169.9	2324.3**	89494.1	87934.0	1560.1
			(1147.9)			(1314.8)

Notes: Means and differences in means (with standard errors in parentheses) of a set of characteristics of school district superintendents and school principals in FP and SP districts (columns 1), and in matched FP and SP districts (columns 2), averaged over the years 2009–2011. The matched sample of SP districts is obtained via nearest-neighbor matching on observable characteristics of the school districts, and includes 56 districts.

Table A3: Propensity Score Matching. Probit, Dependent Variable Equals 1 for FP districts

	(1) β / SE	Marginal effects
enrollment	-0.000	-0.000
	(0.000)	
share disadvantaged students	0.275	0.098
	(0.766)	
salary for teachers w/ exp < 5 (1,000 $)	0.032	0.011
	(0.040)	
salary (1,000 $)	0.050*	0.018*
	(0.029)	
property value per-pupil (1,000 $)	0.000	0.000
	(0.000)	
urban district	0.054	0.019
	(0.467)	
suburban district	-0.044	-0.016
	(0.254)	
expenditure per-pupil (1,000 $)	-0.080**	-0.029**
	(0.039)	
state aid per-pupil (1,000 $)	-0.001	-0.000
	(0.072)	
share of teachers w/ Master degree	-0.364	-0.130
	(0.730)	
share teachers w/ exp ≤ 3	1.535	0.549
	(2.464)	
share teachers w/ exp ≥ 20	-3.278**	-1.172**
	(1.273)	
Observations	224	

Notes: Estimates (column 1) and marginal effects (column 2) of the coefficients of the probit model used to compute the propensity score to match FP districts to SP districts. The dependent variable equals 1 for FP districts. The independent variables are averages of district-level characteristics for the years 2009–2011. *** $p < .01$, ** $p < .05$, * $p < .1$.

Table A4: Summary Statistics, Wisconsin Teachers

	(1) Full sample		(2) FP/SP	
	2007-11	2012-15	2007-11	2012-15
female	0.730	0.742	0.735	0.746
	(0.444)	(0.437)	(0.441)	(0.435)
experience (years)	14.58	13.85	14.36	13.65
	(9.737)	(9.197)	(9.633)	(9.071)
highest ed = BA	0.499	0.469	0.481	0.451
	(0.500)	(0.499)	(0.500)	(0.498)
highest ed = Master	0.494	0.524	0.510	0.541
	(0.500)	(0.499)	(0.500)	(0.498)
highest ed = PhD	0.00179	0.00197	0.00209	0.00231
	(0.0423)	(0.0444)	(0.0456)	(0.0480)
salary ($)	50249.9	53321.2	51167.4	54109.5
	(11327.6)	(12148.0)	(11526.8)	(12364.2)
mover	0.0149	0.0302	0.0138	0.0298
	(0.121)	(0.171)	(0.117)	(0.170)
value-added	-0.0266	0.0327	-0.0211	0.0301
	(0.968)	(1.037)	(0.970)	(1.020)

Notes: Means and standard deviations (in parentheses) of teachers' observable characteristics for the years 2007–2011 and 2012–2015, for all Wisconsin districts (columns 1) and for FP and SP districts (columns 2). The sample only includes teachers for whom VA estimates are available.

Table A5: Summary Statistics, Wisconsin Teachers, With and Without *ex ante* Value-Added

	Full sample			FP/SP		
	w/ *ex ante* VA	w/out *ex ante* VA	Difference	w/ *ex ante* VA	w/out *ex ante* VA	Difference
female	0.78	0.72	0.062∗∗∗	0.79	0.73	0.068∗∗∗
			(0.0014)			(0.0015)
experience (years)	15.3	14.0	1.28∗∗∗	14.9	13.8	1.13∗∗∗
			(0.030)			(0.033)
highest ed = BA	0.42	0.50	-0.084∗∗∗	0.40	0.49	-0.082∗∗∗
			(0.0016)			(0.0017)
highest ed = Master	0.57	0.49	0.085∗∗∗	0.59	0.51	0.083∗∗∗
			(0.0016)			(0.0017)
highest ed = PhD	0.00095	0.0021	-0.0012∗∗∗	0.0012	0.0025	-0.0013∗∗∗
			(0.00014)			(0.00016)
salary ($)	53118.9	51204.3	1914.6∗∗∗	53987.4	52067.5	1919.9∗∗∗
			(37.0)			(41.7)
mover	0.015	0.024	-0.0093∗∗∗	0.015	0.024	-0.0087∗∗∗
			(0.00050)			(0.00055)
value-added	0.00064	-0.0061	0.0068	0.0028	-0.0060	0.0088
			(0.0083)			(0.0092)

Notes: Means and standard deviations (in parentheses) of teachers' observable characteristics for the years 2007–2011 and 2012–2015, for all Wisconsin districts (columns 1) and for FP and SP districts (columns 2), and separately for teachers with and without *ex ante* VA, calculated as the average over the years 2007–2011. The sample includes teachers for whom overall VA estimates are available.

Table A6: Summary Statistics, Districts with and without Handbook Information

	without Handbook	with Handbook	Difference
enrollment	800.8	3027.0	-2226.2***
			(412.6)
disadvantaged students	0.36	0.30	0.057***
			(0.015)
salary for teachers w/ exp $<$ 5 (1,000 $)	35255.5	36442.3	-1186.8***
			(382.5)
teacher salary ($)	48268.4	51493.1	-3224.6***
			(490.4)
math scores (sd)	-0.039	0.12	-0.16***
			(0.055)
teacher experience (yrs)	15.8	15.3	0.54**
			(0.22)
teachers w/ Master	0.40	0.51	-0.11***
			(0.015)
share teachers w/ exp $\leq$ 3	0.12	0.11	0.0069
			(0.0053)
share teachers w/ exp $\geq$ 20	0.32	0.29	0.027***
			(0.0095)
expenditure p.p ($)	16255.6	15346.8	908.8**
			(421.7)
aid p.p ($)	3634.8	4800.3	-1165.5*
			(621.4)

Notes: Means and standard deviations (in parentheses) of district-level characteristics for 102 FP and 122 SP districts with non-missing handbook information, and 203 districts with missing handbook information, for the years 2009–2011.

Table A7: Summary Statistics, Student Data

	(1) Math	(2) Reading
standardized test score	0.00650	0.00494
	(0.997)	(0.998)
share w/disability	0.129	0.129
	(0.335)	(0.335)
share English-learner	0.0504	0.0486
	(0.219)	(0.215)
share free lunch	0.382	0.381
	(0.486)	(0.486)
female	0.489	0.489
	(0.500)	(0.500)
black	0.0959	0.0961
	(0.294)	(0.295)
Hispanic	0.0898	0.0888
	(0.286)	(0.284)
Asian	0.0356	0.0352
	(0.185)	(0.184)
# students in grade-school-year	130.0	129.8
	(96.35)	(96.16)
Observations	3167183	3161997

Notes: Means and standard deviations (in parentheses) of student-level characteristics for the years 2007 to 2016, used to compute teacher VA. The sample includes students in grades 4-8.

Table A8: Teacher Salaries and Value-Added. OLS, Dependent Variable is log(Salary)

	All teachers		Teachers with exp ≤10 yrs	
	(1)	(2)	(3)	(4)
VA	-0.0009	-0.0011	-0.0034	-0.0060*
	(0.0011)	(0.0018)	(0.0026)	(0.0034)
VA * post	0.0012	0.0019	0.0031	0.0054
	(0.0017)	(0.0029)	(0.0037)	(0.0035)
VA * FP * post	0.0034*	0.0030	0.0056	0.0060
	(0.0020)	(0.0026)	(0.0035)	(0.0040)
Year FE	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes
Edu*exp*post	Yes	Yes	Yes	Yes
N	88005	59457	28353	18326
# districts	217	152	216	152

Notes: The dependent variable is the natural logarithm of salaries. The variable *VA* is teacher VA, normalized to have mean 0 and standard deviation 1. The variable *post* equals 1 for years following 2011. All the regressions include year and district fixed effects, as well as indicators for years of experience interacted with indicators for highest education degree interacted with *post*. VA is calculated as the average of a time-varying measure over the years 2007–2011 and 2012–2015. The sample is restricted to tenured teachers (with more than 3 years of experience) working in FP, SP, and matched SP districts, and covers years 2007 to 2015. In columns 3-4, the sample is further restricted to include to teachers with less than 10 years of experience. The matched sample of SP districts is obtained via nearest-neighbor matching on observable characteristics of the school districts. Standard errors in parentheses are clustered at the district level. *** $p < .01$, ** $p < .05$, * $p < .1$.

Table A9: Salaries of Math and Science Teachers. OLS, dependent variable is log(salary)

	FP		SP (full)	
	(1)	(2)	(3)	(4)
math	0.0070		0.0070***	
	(0.0043)		(0.0024)	
math * post	-0.0033		-0.0006	
	(0.0042)		(0.0023)	
VA	-0.0008	-0.0011	-0.0011	-0.0015
	(0.0021)	(0.0023)	(0.0011)	(0.0010)
VA * post	0.0047*	0.0049**	0.0014	0.0020
	(0.0024)	(0.0024)	(0.0015)	(0.0014)
science		0.0101*		0.0053
		(0.0055)		(0.0053)
science * post		-0.0064		-0.0077
		(0.0082)		(0.0073)
Year FE	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes
Edu*exp*post	Yes	Yes	Yes	Yes
Observations	39483	42554	48555	52717
# districts	98	98	119	119

Notes: The dependent variable is the natural logarithm of salaries. The variables *math* and *science* equal 1 for mathematics and science teachers, respectively. The variable *VA* is teacher VA, normalized to have mean 0 and standard deviation equal to 1. The variable *FP* equals 1 for FP districts. The variable *post* equals 1 for years following 2011. All the regressions include year and district fixed effects, as well as indicators for years of experience interacted with indicators for highest education degree interacted with *post*. VA is calculated as the average of a time-varying measure over the years 2007–2011 and 2012–2015. The sample is restricted to tenured teachers (with more than 3 years of experience) working in FP and SP districts, and covers years 2008 to 2015. The matched sample of SP districts is obtained via nearest-neighbor matching on observable characteristics of the school districts. Standard errors in parentheses are clustered at the district level. *** $p < .01$, ** $p < .05$, * $p < .1$.

Table A10: Teacher Sorting. OLS, Dependent Variable Equals 1 for Teachers Moving to or Exiting From a District - Excluding Madison and Milwaukee

	Moving to FP (2008-2014)		Moving to SP (2008-2014)		Exiting from (2008-2012)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	from SP	from FP	from FP	from SP	from FP	from SP (full)	from SP (matched)
high VA	-0.0000	-0.0017	0.0010	0.0010	0.0033	-0.0018	-0.0088
	(0.0014)	(0.0011)	(0.0011)	(0.0010)	(0.0048)	(0.0045)	(0.0058)
high VA * post	0.0038*	0.0014	-0.0049***	-0.0041**	-0.0179*	-0.0051	-0.0093
	(0.0022)	(0.0016)	(0.0018)	(0.0018)	(0.0093)	(0.0081)	(0.0094)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CB controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Teacher controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	26835	30157	30157	26835	17881	16665	8951
# districts	121	100	100	121	98	117	54
Y-mean	0.005	0.004	0.003	0.004	0.041	0.042	0.042

Notes: The dependent variable equals 1 for teachers who move to an FP district (columns 1-2), move to a SP district (column 3-4), exit from an FP district (column 5), exit from a SP district (column 6), and exit from a matched SP district (column 7). In columns 1 and 4 the sample is restricted to teachers already working in a SP district; in columns 2 and 3 it is restricted to those already working in a FP district. The variable *high VA* equals one for teachers with *ex ante* VA above the median. The variable *post* equals 1 for years after Act 10. All the regressions include year and district fixed effects. *District controls* include interactions between 2009–2011 averages of district characteristics interacted with year fixed effects. *CB controls* include an indicator for whether the district had a union recertification election in year t and whether the election was successful. *Teacher controls* include indicators for the number of years of experience and for the highest education degree. Columns 5 and 6 control non-parametrically for age. *Ex ante* VA is calculated as the average of a time-varying measure over the years 2007–2011. The sample is restricted to tenured teachers (with more than 3 years of experience) working in FP and SP districts, and covers years 2008 to 2015 (columns 1-4) and years 2008 to 2012 (columns 5-6). The sample excludes the Milwaukee School District and the Madison Metropolitan School District. Standard errors in parentheses are clustered at the district level. *** $p < .01$, ** $p < .05$, * $p < .1$.

Table A11: Teacher Sorting. OLS, Dependent Variable Equals 1 for Teachers Moving to or Exiting From a District - Matched Sample

	Moving to FP (2008-2014)		Moving to SP (2008-2014)		Exiting from (2008-2012)	
	(1)	(2)	(3)	(4)	(5)	(6)
	from SP	from FP	from FP	from SP	from FP	from SP (full)
high VA	0.0009	-0.0018*	0.0007	-0.0005	0.0033	-0.0088
	(0.0027)	(0.0011)	(0.0008)	(0.0010)	(0.0048)	(0.0058)
high VA * post	0.0020	0.0019	-0.0024*	-0.0004	-0.0179*	-0.0093
	(0.0032)	(0.0015)	(0.0013)	(0.0022)	(0.0093)	(0.0094)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
District controls	Yes	Yes	Yes	Yes	Yes	Yes
CB controls	Yes	Yes	Yes	Yes	Yes	Yes
Teacher controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	14531	30093	30093	14531	17881	8951
# districts	91	100	100	91	98	54
Y-mean	0.008	0.004	0.002	0.004	0.041	0.042

Notes: The dependent variable equals 1 for teachers who move to a FP district (columns 1-2), move to a matched SP district (column 3-4), exit from a FP district (column 5), and exit from a matched SP district (column 6). In columns 1 and 4 the sample is restricted to teachers already working in a SP district; in columns 2 and 3 it is restricted to those already working in a FP district.The variable *high VA* equals one for teachers with *ex ante* VA above the median. The variable *post* equals 1 for years after Act 10. All the regressions include year and district fixed effects. *District controls* include interactions between 2009–2011 averages of district characteristics interacted with year fixed effects. *CB controls* include an indicator for whether the district had a union recertification election in year t and whether the election was successful. *Teacher controls* include indicators for the number of years of experience and for the highest education degree. Columns 5 and 6 control non-parametrically for age. *Ex ante* VA is calculated as the average of a time-varying measure over the years 2007–2011. The sample is restricted to tenured teachers (with more than 3 years of experience) working in FP and matched SP districts, and covers years 2008 to 2015 (columns 1-4) and years 2008 to 2012 (columns 5-6). The matched sample of SP districts is obtained via nearest-neighbor matching on observable characteristics of the school districts. Standard errors in parentheses are clustered at the district level. *** $p < .01$, ** $p < .05$, * $p < .1$.

Table A12: Changes in the Composition of Movers and Exiters. OLS, Dependent Variable is Ex Ante Teacher Value-Added - Matched Sample

	(1) Movers	(2) Movers	(3) Exiters	(4) Exiters
FP	0.0995	0.1058	0.0878	0.0303
	(0.2097)	(0.2137)	(0.0918)	(0.1003)
FP * post	0.2665	0.3113	-0.2831***	-0.2555***
	(0.3149)	(0.3652)	(0.0895)	(0.0911)
Year FE	Yes	Yes	Yes	Yes
District controls	Yes	Yes	Yes	Yes
CB controls	Yes	Yes	Yes	Yes
Teacher controls	Yes	Yes	Yes	Yes
Budget controls	No	Yes	No	Yes
Observations	486	484	1703	1406
# districts	111	109	151	146

Notes: The dependent variable is *ex ante* teacher VA. Columns 1-2 and 5-6 are estimated on the subsample of movers to a district; columns 3-4 are estimated on the subsample of leavers from a district (defined as teachers who leave Wisconsin's teaching workforce). The variable *FP* equals 1 for FP districts. The variable *post* equals 1 for years following 2011. All the regressions include year fixed effects. *District controls* include interactions between the 2009–2011 averages of district characteristics interacted with year fixed effects. *CB controls* include an indicator for whether the district had a union recertification election in year t and whether the election was successful. *Teacher controls* include indicators for the number of years of experience and for the highest education degree. *Budget controls* are district-year-level controls for the level of state aid as a share of total revenues, as well as per-teacher expenditure on salaries, retirement, health, life, and other insurance, and other employee benefits. Columns 1-2 include indicators for the type district of origin (FP or SP), interacted with *FP* and with *post*. Columns 3-4 control non-parametrically for age. Ex ante VA is calculated as the average of a time-varying measure over the years 2007–2011. The sample is restricted to tenured teachers (with more than 3 years of experience) working in FP and matched SP districts, and covers years 2008 to 2015. The matched sample of SP districts is obtained via nearest-neighbor matching on observable characteristics of the school districts. Standard errors in parentheses are clustered at the district level. *** $p < .01$, ** $p < .05$, * $p < .1$.

Table A13: Changes in the Composition of the Teaching Workforce. OLS, Dependent Variable is Ex Ante Teacher Value-Added - Excluding Madison and Milwaukee

	Full sample				Matched sample			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
FP	-0.0325	-0.0341	-0.0278		-0.0507	-0.0526	-0.0705	
	(0.0856)	(0.0855)	(0.0825)		(0.0955)	(0.0951)	(0.0883)	
FP * post	0.0417	0.0438	0.0611**	0.0222*	0.0467	0.0473	0.0804**	0.0410**
	(0.0266)	(0.0267)	(0.0304)	(0.0130)	(0.0301)	(0.0301)	(0.0398)	(0.0159)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CB controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Teacher controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Budget controls	No	No	Yes	Yes	No	No	Yes	Yes
District FE	No	No	No	Yes	No	No	No	Yes
Observations	58600	58524	58224	58224	45901	45825	45525	45525
# districts	215	215	212	212	152	152	149	149

Notes: The dependent variable is *ex ante* teacher VA. The variable *FP* equals 1 for FP districts. The variable *post* equals 1 for years following 2011. All the regressions include year fixed effects. *District controls* include interactions between the 2009–2011 averages of district characteristics interacted with year fixed effects. *CB controls* include an indicator for whether the district had a union recertification election in year *t* and whether the election was successful. *Teacher controls* include indicators for the number of years of experience and for the highest education degree. *Budget controls* are district-year-level controls for the level of state aid as a share of total revenues, as well as per-teacher expenditure on salaries, retirement, health, life, and other insurance, and other employee benefits. Ex ante VA is calculated as the average of a time-varying measure over the years 2007–2011. The sample is restricted to tenured teachers (with more than 3 years of experience) working in FP, SP, and matched SP districts, excludes the Milwaukee School District and the Madison Metropolitan School District, and covers years 2008 to 2015. The matched sample of SP districts is obtained via nearest-neighbor matching on observable characteristics of the school districts. Standard errors in parentheses are clustered at the district level. *** $p < .01$, ** $p < .05$, * $p < .1$.

Table A14: Changes in Teacher Effort. OLS and 2SLS, Dependent Variable is Teacher Value-Added - Excluding Madison and Milwaukee

	Full sample			Matched sample		
	(1) Selection + Effort	(2) Effort (Incumbents)	(3) Effort (IV)	(4) Selection + Effort	(5) Effort (Incumbents)	(6) Effort (IV)
FP	-0.0278	-0.0108	-0.0269	-0.0750	-0.1759*	-0.0760
	(0.0809)	(0.0854)	(0.0812)	(0.0872)	(0.0978)	(0.0879)
FP * post	0.0854	0.0768	0.0737	0.1485*	0.1759**	0.1308*
	(0.0617)	(0.0659)	(0.0608)	(0.0784)	(0.0730)	(0.0757)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
District controls	Yes	Yes	Yes	Yes	Yes	Yes
CB controls	Yes	Yes	Yes	Yes	Yes	Yes
Teacher controls	Yes	Yes	Yes	Yes	Yes	Yes
Budget controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	57388	42024	56847	44858	32765	44392
# districts	212	212	212	149	149	149

Notes: The dependent variable is VA of all teachers (columns 1, 3, 4, 6) and incumbent teachers (columns 2 and 5). Incumbent teachers are defined as those who do not change district nor exit Wisconsin public schools after Act 10. The variable *FP* equals 1 for FP districts. In columns 3 and 6, the variable *FP* is instrumented with an indicator for whether a teacher has taught at least once in a FP district between 2007 and 2011. The variable *post* equals 1 for years following 2011. All the regressions include year fixed effects. *District controls* include interactions between 2009–2011 averages of district characteristics and year dummies. *CB controls* include an indicator for whether the district had a union recertification election in year t and whether the election was successful. *Teacher controls* include indicators for each number of years of experience and for the highest education title (bachelor, Master, Ph.D.). VA is calculated as the average of a time-varying measure over the years 2007–2011 and 2012–2015.The sample is restricted to tenured teachers (with more than 3 years of experience) working in FP, SP, and matched SP districts, excludes the Milwaukee School District and the Madison Metropolitan School District, and covers years 2008 to 2015. The matched sample of SP districts is obtained via nearest-neighbor matching on observable characteristics of the school districts. Standard errors in parentheses are clustered at the district level. *** $p < .01$, ** $p < .05$, * $p < .1$.

Table A15: Changes in Student Characteristics, Movers to FP and SP

	(1) Share low SES	(2) share Black	(3) share Hispanic	(4) share EL status
FP * post	0.0410*	-0.0028	0.0089	0.0018
	(0.0242)	(0.0053)	(0.0099)	(0.0053)
School FE	Yes	No	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	529	529	529	529
# districts	114	114	114	114

Notes: The dependent variables are average characteristics of students of the teachers who move to FP and SP districts. The variable *FP* equals 1 for 102 FP districts, the variable *post* equals 1 for years after 2011. All specifications contain district and year fixed effects. Standard errors in parentheses are clustered at the district level.

Table A16: Sample for Model Estimation - Summary Statistics, 2014

	(1)
Teachers (N = 12573)	
female	0.7920
	(0.4059)
has BA	0.4106
	(0.4920)
has Master	0.5843
	(0.4929)
has PhD	0.0009551
	(0.03089)
experience (years)	15.358
	(8.8115)
value-added	0.03854
	(1.0309)
salary ($)	55380.3
	(11814.5)
moves district	0.01656
	(0.1276)
exits	0.1172
	(0.3216)
distance, movers (miles)	19.659
	19.197
Districts (N = 410)	
enrollment	1949.3
	4178.4
share disadvantaged students	0.3829
	0.1547
urban	0.03922
	0.1943
suburban	0.1495
	0.3570

Notes: Means and differences in means (with standard errors in parentheses) of characteristics of teachers and districts included in the estimation of the model.

Appendix B Estimating Teacher Value-Added With Grade-School Links

Teacher value-added is defined as the contribution of each teacher to achievement (or achievement growth), once all other determinants of student learning have been taken into account. The starting model is the following (Chetty et al., 2014a):

$$\begin{aligned} A^*_{kt} &= \beta X_{kt} + \nu_{kt} \\ \text{where } \nu_{kt} &= \mu_{i(kt)} + \theta_{c(it)} + \varepsilon_{kt} \end{aligned} \tag{19}$$

and where A^*_{kt} a standardized measure of test scores (or test score gains) for student k in year t, and X_{kt} is a vector of student observables which could affect achievement (such as demographic characteristics and past test scores) as well as school characteristics.[60] The residual ν_{kt} can be decomposed in three parts: The error term component $\mu_{i(kt)}$ is VA of teacher i, teaching student k in year t; the component $\theta_{c(kt)}$ is an exogenous classroom shock, and ε_{kt} is an idiosyncratic student-specific component which varies over time. VA estimates are estimates of the teacher effect μ_i.

A range of techniques have been proposed to estimate μ_i, including fixed effects (Aaronson et al., 2007) and two-steps procedures based on the decomposition of test score residuals (Kane and Staiger, 2008; Chetty et al., 2014a). Here, I consider the two-steps estimator of Chetty et al. (2014a), which allows for drifts in teacher quality over time, and corrects teacher effects for noise in the estimates using a Bayes approach. Other estimators proposed by the literature (such as Kane and Staiger, 2008) can be seen as special cases of this method.

The estimation procedure can be summarized as follows:

1. Regress A^*_{kt} on X_{kt} and estimate β via OLS;
2. Construct residuals $A_{kt} = A^*_{kt} - \hat{\beta} X_{kt}$, where $\hat{\beta}$ is the OLS estimate of β;
3. Estimate $\hat{\mu}_{it}$ as $\hat{\mu}_{it} = \sum_{m=t-l}^{t-1} \psi_m \bar{A}_{im}$, where $\bar{A}_{it}$ are average test score residuals of all students taught by teacher i in year t, l is the number of lags, and

$$\psi = \arg \min_{\{\psi_1,\ldots,\psi_{t-1}\}} \sum_i \left(\bar{A}_{it} - \sum_{s=1}^{t-1} \psi_s \bar{A}_{is} \right)^2$$

Clearly, calculating $\bar{A}_{it}$ requires matching each teacher with the students she taught.[61] The WDPI started to record classroom identifiers (necessary to perform these links) only in 2017-18; data from previous years only contain identifiers for schools and grades. This means that, in a given year, a student can be linked to all the teachers in his school and grade, but not to the specific teacher who taught him (and conversely, a teacher can be linked to all students attending her grade in her school, but not to her own pupils). The lack of information on classroom identifiers is common to teacher-student datasets from several other states and/or districts (Rivkin et al., 2005, for example, face a similar issue with data from Texas).

How to identify teacher effects in the absence of classroom links? Rivkin et al. (2005) show how, in the presence of teacher turnover across grades or schools, one can use grade-school level residuals to obtain a more accurate measure of teacher effects than the simple grade average. To incorporate this feature of

[60] Several types of achievement models have been used in the literature, including specifications with student fixed effects and/or with the student's prior test scores.

[61] Generally, elementary-school teachers (grades 1 through 5) teach all subjects in only one classroom, whereas middle-school teachers teach one subject in multiple classrooms.

the data, the estimator can be modified as follows:

$$\hat{\mu}_{it} = \sum_{m=t-l}^{t-1} \psi_m \bar{A}^g_{im}$$

$$\text{where} \quad \bar{A}^g_{it} = \frac{1}{N_{gst}} \sum_{k:g^e(it)=g^p(kt),s^e(it)=s^p(kt)} A_{kt} \text{ and } A_{kt} = A^*_{kt} - \hat{\beta} X_{kt}$$

where $g^e(it)$ ($s^e(it)$) is the grade (school) where teacher i teaches in year t, and $g^p(kt)$ ($s^p(kt)$) is the grade (school) attended by student k in year t. The quantity A^g_{it} represents the average test score residuals for students in grade g, school s, and year t.

The intuition behind the identification of teacher effects follows Rivkin et al. (2005). In the absence of teacher turnover, teachers in grade g and school s would have the same $\bar{A}^g_{it}$ for every t, and would be assigned the same estimate (equal to the average of their true effects). With data on test scores for multiple years and in the presence of turnover, teachers switches across schools or within schools and grades allow to isolate the effect of the individual teacher through the comparison of test score residuals before and after her arrival in a given grade and school. Importantly, teacher turnover allows a more precise identification of the effects not only of the teacher who switches school or grade, but also of the teachers teaching in her same grade and school at any point in time.

Under the assumptions that μ_{it} and ε_{it} follow stationary processes (i.e. $\mathbb{E}(\mu_{it}|t) = \mathbb{E}(\varepsilon_{it}|t) = 0$, $Cov(\mu_{it}, \mu_{it+s}|t) = \sigma_{\mu s}$, and $Cov(\varepsilon_{it}, \varepsilon_{it+s}|t) = \sigma_{\varepsilon s}$), this estimator is unbiased. Two features of this identification strategy are worth highlighting:

1. While these estimates of teacher VA are more precise than grade-school residuals, they will still contain more noise relative to estimates obtained with links. Even in the presence of turnover, teachers always teaching the same grade-school would have the same $\bar{A}^g_{it}$ for every t, and hence the same estimate.

2. The aggregation of teacher effects at the grade level overcomes a problematic form of selection, which occurs within schools and grades and across classrooms when some parents manage to have their children assigned to specific teachers. The (forced) use of grade-school estimates circumvents this form of selection, and is in practice equivalent to an instrumental variable estimator based on grade rather than classroom assignment (Rivkin et al., 2005).

Appendix B.1 Identification of Teacher Value-Added With Turnover

To understand the identification argument, consider a simple example of 3 teachers (A, B, C) observed in 3 periods ($t = 1, 2, 3$) and in 2 possible grades ($g = 4, 5$). The teaching assignments are as follows.

period	grade	
1	A,B	C
2	B,C	A
3	A,C	B

The objective is to calculate VA of the three teachers in period 3. I define A_{kt} as the average test score residual for students of teacher k in period t, and $\bar{A}^g_t$ the average test score residuals of students in grade g in period t. Following Chetty et al. (2014a) I can write the VA estimate for each teacher as follows (I suppress the hats on the VA estimates for ease of notation and I consider 3 lags):

$$\mu_{A3} = \begin{bmatrix} A_{A1}^2 & A_{A1}A_{A2} \\ A_{A1}A_{A2} & A_{A2}^2 \end{bmatrix}^{-1} \begin{bmatrix} A_{A1}A_{A3} \\ A_{A2}A_{A3} \end{bmatrix} \quad (20)$$

$$\mu_{B3} = \begin{bmatrix} A_{B1}^2 & A_{B1}A_{B2} \\ A_{B1}A_{B2} & A_{B2}^2 \end{bmatrix}^{-1} \begin{bmatrix} A_{B1}A_{B3} \\ A_{B2}A_{B3} \end{bmatrix} \quad (21)$$

$$\mu_{C3} = \begin{bmatrix} A_{C1}^2 & A_{C1}A_{C2} \\ A_{C1}A_{C2} & A_{C2}^2 \end{bmatrix}^{-1} \begin{bmatrix} A_{C1}A_{C3} \\ A_{C2}A_{C3} \end{bmatrix} \quad (22)$$

Assuming a constant number of students in each classroom, one can write:

$$\bar{A}_1^4 = \frac{1}{2}(A_{A1} + A_{B1}) \quad (23)$$

$$\bar{A}_1^5 = A_{C2} \quad (24)$$

$$\bar{A}_2^4 = \frac{1}{2}(A_{B2} + A_{C2}) \quad (25)$$

$$\bar{A}_2^5 = A_{A2} \quad (26)$$

$$\bar{A}_3^4 = \frac{1}{2}(A_{A3} + A_{C3}) \quad (27)$$

$$\bar{A}_3^5 = A_{B3} \quad (28)$$

My VA estimator implies:

$$\mu_{A3} = \begin{bmatrix} (\bar{A}_1^4)^2 & \bar{A}_1^4\bar{A}_2^5 \\ \bar{A}_1^4\bar{A}_2^5 & (\bar{A}_2^5)^2 \end{bmatrix}^{-1} \begin{bmatrix} \bar{A}_1^4\bar{A}_3^4 \\ \bar{A}_2^5\bar{A}_3^4 \end{bmatrix} \quad (29)$$

$$\mu_{B3} = \begin{bmatrix} (\bar{A}_1^4)^2 & \bar{A}_1^4\bar{A}_2^4 \\ \bar{A}_1^4\bar{A}_2^4 & (\bar{A}_2^4)^2 \end{bmatrix}^{-1} \begin{bmatrix} \bar{A}_1^4\bar{A}_3^5 \\ \bar{A}_2^4\bar{A}_3^5 \end{bmatrix} \quad (30)$$

$$\mu_{C3} = \begin{bmatrix} (\bar{A}_1^5)^2 & \bar{A}_1^5\bar{A}_2^4 \\ \bar{A}_1^5\bar{A}_2^4 & (\bar{A}_2^4)^2 \end{bmatrix}^{-1} \begin{bmatrix} \bar{A}_1^5\bar{A}_3^4 \\ A_{C2}\bar{A}_3^4 \end{bmatrix} \quad (31)$$

Equations (20)-(31) represent a system of 12 equations in 12 unknowns: μ_{A3}, μ_{B3}, μ_{B3}, A_{A1}, A_{A2}, A_{A3}, A_{B1}, A_{B2}, A_{B3}, A_{C1}, A_{C2}, A_{C3}. In this case, VA of teachers can be perfectly identified because at least one teacher switches grade each year.

Appendix B.2 Validation Exercise: Value-Added with Classroom Links and with Grade-School Links in the NYC data

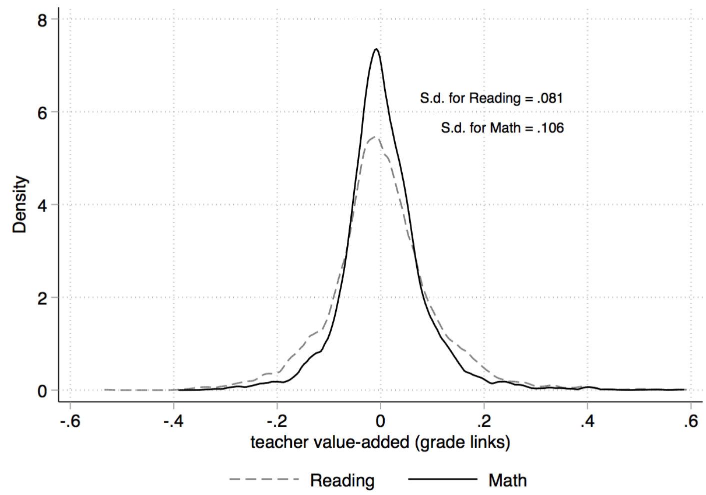
To validate the VA estimator with grade-school links (GS) against the standard estimator with classroom links (CL), I use teacher and student data from the New York City Department of Education (NYCDOE) from the years 2006-07 to 2009-10. This dataset contains classroom, grade, and school identifiers, which allow me to estimate both CL and GL measures. I estimate teacher VA for 15,469 teachers of math and English-Language-Arts (ELA) using the procedure of Chetty et al. (2014a).

Measurement Error. The main limitation of GL estimates relative to CL is measurement error. Since students are linked to teachers at the grade-school level, VA of a teacher will also be a function of test scores of students she never taught.

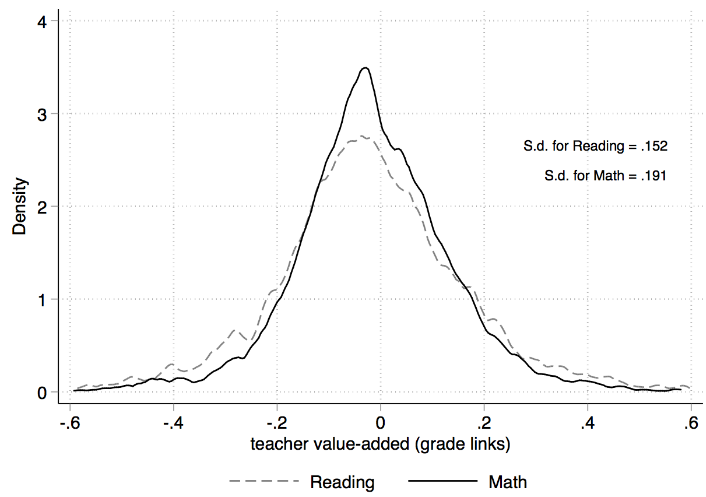
To quantify the degree of measurement error, Figure B1 shows the kernel density of the distribution of GL (top panel) and CL (bottom panel). As expected, the distribution of GL is more concentrated around zero compared to CL. In spite of this, GL is able to explain a significant amount of variance in test scores. Its standard deviation (measured in test scores standard deviation units) is equal to 0.11 for math teachers; by comparison, the standard deviation of CL is equal to 0.19. Figure B2 shows the density of GL for Wisconsin teachers. Its standard deviation is equal to 0.08 for math teachers.

Figure B1: Empirical Distribution of Value-Added Estimates: New York City, 2007-2010

Panel A: Value-Added with Grade-School Links



Panel B: Value-Added with Classroom Links



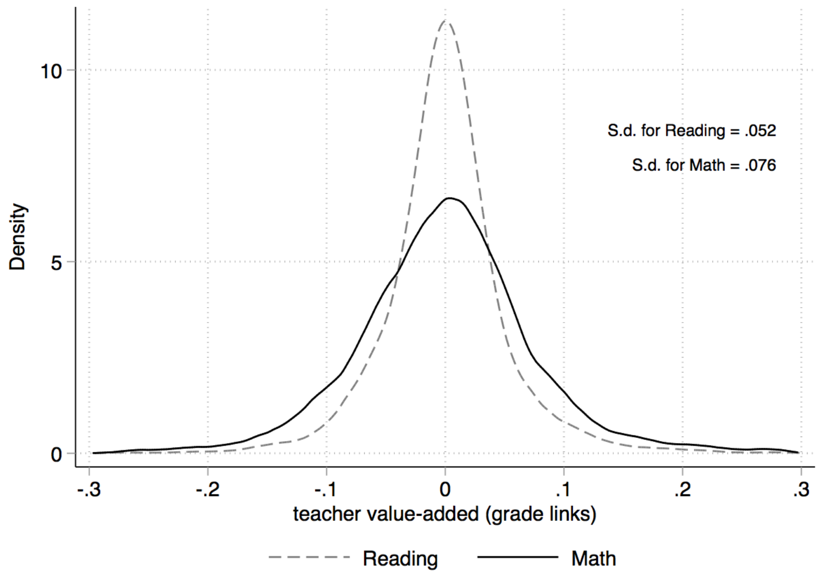
Notes: Kernel densities of the empirical distribution of VA estimates for NYC math and ELA teachers, for each subject. Estimates are averaged across years for each teacher. Each density is weighted by the number of student test scores observations used to estimate each teacher's VA, and estimated using a bandwidth of 0.05. The figure also reports the standard deviations of these empirical distributions.

Forecast Bias of GL as a Proxy for CL. I assess whether GL is a forecast-unbiased estimate for CL. Figure B3 shows a binned scatterplot of the two estimates in the NYC data, averaged across the four years for each teacher. Their correlation is 0.62. The forecast bias of $\hat{\mu}_i^{GL}$ as a proxy for $\hat{\mu}_i^{CL}$ can be defined based on the best linear predictor of $\hat{\mu}_i^{CL}$ given $\hat{\mu}_i^{GL}$:

$$\hat{\mu}_i^{CL} = \alpha + \gamma \hat{\mu}_i^{GL} + \chi_i \tag{32}$$

Assuming χ_i to be uncorrelated with $\hat{\mu}_i^{GL}$, the forecast bias f is zero if $\gamma = 1$: $f = 1 - \gamma$. I can estimate the slope coefficient γ estimating Equation (32) via OLS. 95% confidence intervals of γ, whose point estimate

Figure B2: Empirical Distribution of Value-Added Estimates: Wisconsin, 2007-2015



Notes: Kernel densities of the empirical distribution of VA estimates for Wisconsin math and reading teachers, for each subject. Estimates are averaged across years for each teacher, separately for years before and after Act 10. Each density is weighted by the number of student test scores observations used to estimate each teacher's VA, and estimated using a bandwidth of 0.05. The figure also reports the standard deviations of these empirical distributions.

is equal to 0.94, include 1, which implies that the forecast bias f is indistinguishable from zero (Figure B3).

Teacher Switches as a Quasi-Experiment. As an additional test for the unbiasedness of GL estimates I exploit teacher switches across grades as a quasi-experiment, as done by Chetty et al. (2014a). If VA is an unbiased measure of teacher quality, changes in average VA of teachers in a given school and grade (driven by teacher switches) should predict changes in average student test score residuals one-by-one. To understand the rationale behind this test suppose that, in a given school with three 4th-grade classrooms (and hence three 4th-grade math teachers), one of these teachers leaves and is replaced by a teacher with a 0.3 higher VA (measured in standard deviations of test scores). If VA is an unbiased measure of teacher effectiveness, test scores should raise by 0.3/3 = 0.1 due to this switch (Chetty et al., 2014a).

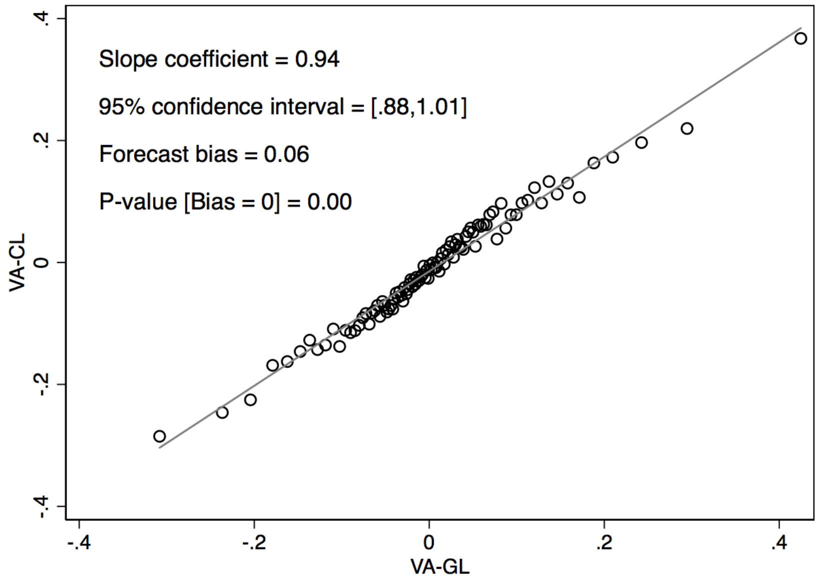
I estimate the degree of forecast bias for the Wisconsin GL measures by estimating the following first-differences equation:

$$\Delta A_{gst} = a + b\Delta Q_{gst} + \Delta\chi_{gst} \tag{33}$$

where Q_{gst} is average VA of teachers in grade g, school s, and year t and where $\Delta W_{gst} = W_{gst} - W_{gst-1}$. The bias is then defined as $\lambda = 1 - b$. Table B3 shows estimates of b and λ, obtained using either mean residual test scores A_{gst} or mean actual test scores A^*_{gst}, and controlling also for school-by-year fixed effects (as done by Chetty et al., 2014a).[62] Estimates of b are all close to 1 both over the full sample period and in the years after Act 10. While slightly larger than Chetty et al. (2014a), who estimate it to be between 0.003 and 0.026, estimates of b are all close to 0.05 and indistinguishable from zero, both over the full sample period and in the years after Act 10.

[62]The fact that using A^*_{gst} as a regressor instead of A_{gst} yields similar results further confirms that selection of students across teachers is unlikely to generate substantial bias in the estimates (Chetty et al., 2014a).

Figure B3: Binned scatterplot: $\hat{\mu}_i^{CL}$ and $\hat{\mu}_i^{GL}$



Notes: The figure shows the relationship between $\hat{\mu}_i^{CL}$, estimate of teacher VA obtained using the procedure of Chetty et al. (2014a) and teacher-student links, and $\hat{\mu}_{GL}$, its analogous obtained discarding these links. Estimates are obtained using data from New York City students and teachers of math and ELA, for the years 2007-2010.

Table B1: Forecast bias in teacher VA

	$\Delta\bar{A}_{gst}$		$\Delta\bar{A}^*_{gst}$	
	(1) full sample	(2) post-Act 10	(3) full sample	(4) post-Act 10
ΔVA_{gst}	0.977	1.048	1.048	1.073
	(0.155)	(0.223)	(0.220)	(0.285)
School-by-year FE	Yes	Yes	Yes	Yes
Observations	32849	14668	32849	14668
# districts	418	414	418	414
λ	0.023	-0.048	-0.048	-0.073
p-value λ=0	0.88	0.83	0.83	0.80

Notes: The dependent variable is the first difference in grade-school average test score residuals (from a regression of test scores on student characteristics, school, and grade fixed effects, columns 1 and 2) or in average test scores at the grade, school, and year level (columns 3 and 4). The variable ΔVA_{gst} is the first difference in average teacher VA in school s and grade g. VA is calculated using data from Wisconsin, and is averaged over the years 2007-2011 and 2012-2015. All regressions include school-by-year fixed effects. Standard errors in parentheses are clustered at the district level.

Non-Classical Measurement Error. A possible concern with VA-GL is non-classical measurement error, which occurs when the precision of the estimates is related to characteristics of the teachers or the students. This issue could arise, for example, if teachers who switch across schools or grades (and, analogously, the grades and schools employing these teachers) are selected on the basis of observable and/or unobservable characteristics.

In Table B1 I to use the GL and CL estimates of VA from the NYC data to investigate the extent of measurement error. Specifically, I correlate the difference between GL and CL (a proxy for measurement error) with a range of student and teacher observable characteristics. These estimates reveal no discernible relationship between the error and these characteristics, with the exception of the share of special education students. Importantly, the measurement error does not appear to be systematically different between teachers who switch across grades and teachers who do not. While this evidence does not guarantee that the error is not related to unobservables, it reassuringly shows no systematic patterns of correlations between VA and this set of student and teacher observables.

Table B2: Correlations Between the Difference [GL-CL] and Student and Teacher Observables

	Math	ELA
	(1)	(2)
experience	0.0001	-0.0007
	(0.0005)	(0.0005)
switcher	0.0007	-0.0023
	(0.0046)	(0.0038)
Black	-0.0093	-0.0004
	(0.0063)	(0.0053)
Asian	-0.0003	0.0160
	(0.0091)	(0.0102)
Hispanic	-0.0003	0.0015
	(0.0077)	(0.0063)
% low SES students	0.0014	0.0059
	(0.0155)	(0.0137)
% Black students	0.0207	0.0063
	(0.0147)	(0.0135)
% Hispanic students	0.0255	0.0163
	(0.0166)	(0.0146)
% Asian students	0.0438**	0.0084
	(0.0201)	(0.0167)
% special Ed students	-0.1471***	-0.1269***
	(0.0077)	(0.0069)
Observations	11109	11497

Notes: OLS regression of the difference between GL and CL and a range of student and teacher characteristics, averaged at the teacher-year level. VA is calculated using data from NYC. Robust standard errors in parentheses.

Appendix C Roy Model

To provide a conceptual framework for the effects of changes in pay on teacher sorting across districts, I build a model similar to Roy (1951). I consider two equally-sized districts, A and B, and a continuum of utility-maximizing teachers, having utility $u = w$ (where w is salary) and a level of quality θ, normally distributed with mean μ and variance σ^2.

In each of two periods ($t = 0$ and $t = 1$) teachers decide where to teach. At time $t = 0$, salaries are fixed in all districts and do not depend on teacher quality: $w_A = w_B = \bar{w}$. Since districts are identical, teachers equally distribute themselves between A and B, and the resulting distribution of teachers' quality is identical across the two districts. Under these circumstances teachers have no incentives to change district, and the moving rate is therefore zero. I consider a change in salaries in A at time $t = 1$, from $\bar{w}$ to $w_A = \alpha\bar{w} + \beta\theta$, with $0 < \alpha < 1$ and $0 < \beta < 1$. Salaries in B remain unchanged at $\bar{w}$.

Two simple propositions summarize the implications of the salary change in district A on the sorting patterns of teachers across districts in $t = 1$.

Proposition 1. The share of teachers moving across districts increases after a change in salary schemes in one of the two districts.

Proof. The share of teachers moving at $t = 0$ is zero. Utility maximization implies that a teacher with quality θ will move from district A to district B if and only if $w_A < w_B$.[63] As a result, the share of teachers who move from A to B is:

$$s_{AB} = Pr(\alpha\bar{w} + \beta\theta < \bar{w}) = \Phi\left(\frac{(1-\alpha)\bar{w} - \beta\mu}{\beta\sigma}\right) \tag{34}$$

where $\Phi(.)$ denotes the CDF of a standard normal. Analogously, the share of teachers who move from B to A is:

$$s_{BA} = Pr(\text{move from B to A}) = P(\alpha\bar{w} + \beta\theta > \bar{w}) = 1 - \Phi\left(\frac{(1-\alpha)\bar{w} - \beta\mu}{\beta\sigma}\right) \tag{35}$$

As long as $\beta > 0$, $s_{AB} > 0$ and $s_{BA} > 0$. □

Proposition 2. Movers from A to B are negatively selected in the district of origin. Movers from B to A are positively selected in the district of origin.

Proof. Define $k = (1-\alpha)w/\beta > 0$ and $\kappa = (k-\mu)/\sigma$. By the properties of a truncated normal, the average quality of teachers moving from A to B is

$$\widetilde{\theta}_{AB} = \mathbb{E}_\theta(\theta|\alpha\bar{w} + \beta\theta < \bar{w}) = \mu - \sigma\frac{\phi(\kappa)}{\Phi(\kappa)} \tag{36}$$

where $\phi(.)$ is the PDF of a standard normal. Similarly, the average quality of teachers moving from B to A is

$$\widetilde{\theta}_{BA} = \mathbb{E}_\theta(\theta|\alpha\bar{w} + \beta\theta > \bar{w}) = \mu + \sigma\frac{\phi(\kappa)}{1-\Phi(\kappa)} \tag{37}$$

It is trivial to demonstrate that $\widetilde{\theta}_{AB} < \mu$ and $\widetilde{\theta}_{BA} > \mu$. The result follows. □

Intuitively, the introduction of a quality component of teacher salaries in one district changes the optimal choice of some teachers, leading to an increase in cross-district movements. This happens because lower-quality teachers are over-compensated and higher-quality teachers are under-compensated in a fixed-salary regime (t=0), compared to a case in which quality is rewarded in district A (t=1). As a result, in the second period higher-quality teachers working in B will be better off moving to A in order to increase their salary, whereas lower-quality teachers in A will be better off moving to B in order to maintain their original salary level.

[63] I assume, without loss of generality, that moving costs are zero for all teachers.

This simple framework can be extended to account for exit from the teaching profession, by allowing for the existence of an outside labor market, denoted by O. This market represents the outside option for teachers currently employed in one of the two districts, such as a teaching job in a private schools, in a different sector, non-employment, or retirement. When employed in O, workers receive a salary $w_O = \gamma\bar{w} + \lambda\theta$. To rule out uninteresting cases, I assume that exit entails no cost, whereas moving entails a cost m. The presence of the outside market guarantees that teachers who remain in public schools at $t = 0$ have quality $\theta < \bar{\theta}$, where $\bar{\theta} = \frac{(1-\gamma)\bar{w}}{\lambda}$. This setup implies the following propositions.

Proposition 3. Under the above assumptions, the share of teachers who exit public schools increases after a change in the salary scheme in one of the two districts.

Proof. The share of teachers exiting at $t = 0$ is zero. Utility maximization implies that a teacher with quality θ working in A will exit if and only if $w_O > \max\{w_A, w_B - m\}$. Similarly, a teacher with quality θ working in B will exit if and only if $w_O > \max\{w_A - m, w_B\}$. Since all teachers working in districts A and B have quality $\theta < \bar{\theta}$, the share of teachers exiting from B in $t = 1$, s_{BO}, is equal to 0.

To calculate the share of teachers exiting from A, s_{AO}, I define some useful quantities. I define θ_1 as the quality of a teacher currently working in A who is indifferent between exiting and moving to B. Similarly, I define θ_2 as the quality of a teacher currently working in A who is indifferent between exiting and remaining in A.[64] Lastly, I define θ_3 as the quality of a teacher working in A who is indifferent between staying in A and moving to B.

$$\theta_1 = \frac{(1-\gamma)\bar{w} - m}{\lambda} \tag{38}$$

$$\theta_2 = \frac{(\gamma-\alpha)\bar{w}}{\beta - \lambda} \tag{39}$$

$$\theta_3 = \frac{(1-\alpha)\bar{w} - m}{\beta} \tag{40}$$

To rule out uninteresting cases, I assume $\theta_2 < \bar{\theta}$ and $\theta_2 > \max\{\theta_1, \theta_3\}$, i.e. there always exist teachers who prefer to exit A, and teachers who prefer to stay. Under these assumptions, the following cases can occur:

- $\theta_1 < \theta_3$. In this case, described in the top panel of Figure C1, three groups of teachers exists among the incumbents in A: those with $\theta < \theta_1$ (red section), who will be better off moving to B; those with $\theta : \theta_1 < \theta < \theta_2$ (green section), who will be better off exiting; and those with $\theta > \theta_2$, who will be better off staying (blue section).
- $\theta_3 < \theta_1$. In this case, described in the bottom panel of Figure C1, teachers with $\theta < \theta_3$ (red section) will be better off moving to B; those with $\theta : \theta_3 < \theta < \theta_2$ (blue section) will be better off staying in A; and those with $\theta > \theta_2$ (green section) will be better off exiting.

  As a result, s_{AO} can be written as:

$$s_{AO} = \begin{cases} Pr(\theta_1 < \theta < \theta_2) = \Phi(\frac{\theta_2-\mu}{\sigma}) - \Phi(\frac{\theta_1-\mu}{\sigma}) & \text{if } \theta_1 < \theta_3 \\ Pr(\theta > \theta_2) = 1 - \Phi(\frac{\theta_2-\mu}{\sigma}) & \text{if } \theta_3 < \theta_1 \end{cases} \tag{41}$$

  which is positive under the above assumptions.

□**Prediction 4**. Under the above assumptions, exiters from A are negatively selected in the district of origin.

Proof. Define $\rho = (\bar{\theta} - \mu)/\sigma$, $\upsilon = (\theta_1 - \mu)/\sigma$, and $\varphi = (\theta_2 - \mu)/\sigma$. Note first that the average quality of teachers in A and B before the change in the salary scheme is $\mathbb{E}_\theta(\theta|\theta < \bar{\theta}) = \mu - \sigma\phi(\rho)/\Phi(\rho) < \mu$. By the

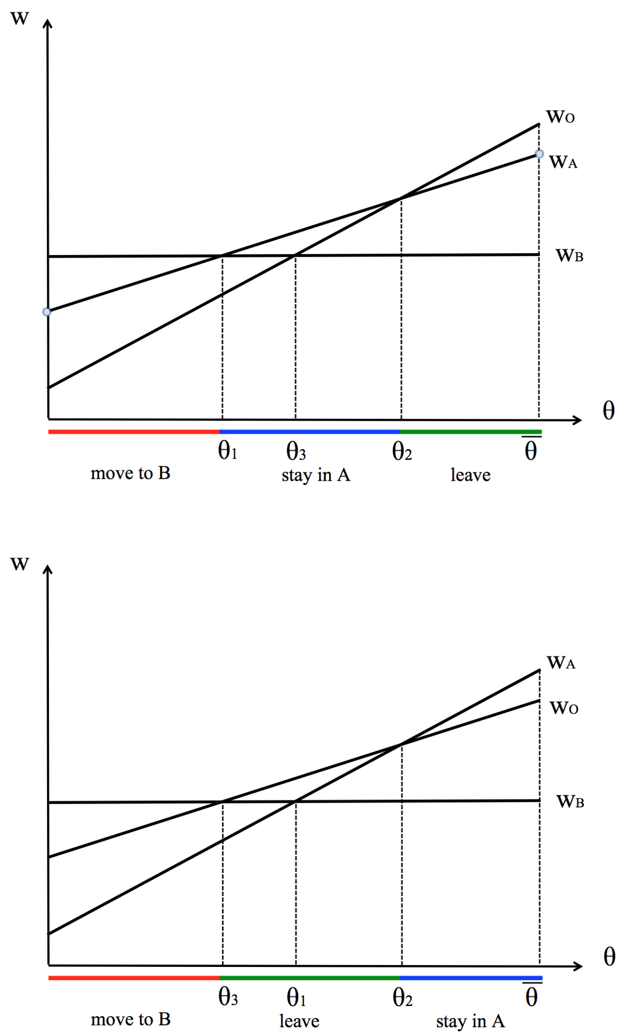
[64] I assume θ_2 exists and is finite.

properties of a truncated normal, the average quality of teachers exiting from A is

$$\widetilde{\theta}_{AO} = \begin{cases} \mathbb{E}_\theta(\theta|\theta_1 < \theta < \theta_2) = \mu - \sigma \frac{\phi(\varphi) - \phi(\upsilon)}{\Phi(\varphi) - \Phi(\upsilon)} & \text{if } \theta_1 < \theta_3 \\ \mathbb{E}_\theta(\theta|\theta > \theta_2) = \mu + \sigma \frac{\phi(\varphi)}{1 - \Phi(\varphi)} & \text{if } \theta_3 < \theta_1 \end{cases} \tag{42}$$

If $\phi(\rho)/\Phi(\rho) < (\phi(\varphi) - \phi(\upsilon))/(\Phi(\varphi) - \Phi(\upsilon))$, the first case implies the result. □

Figure C1: Conceptual framework: movers and leavers, case 1 (top panel) and case 2 (bottom panel)



Notes: Values of θ and move/exit decisions for incumbents in district A, as described in the conceptual framework in Section 5.

Appendix D Solving the And Estimating the Model

Solving the District's Problem

District j faces the following problem (I drop the subscript j for simplicity):

$$\begin{aligned} & \max_{\{\mathbf{o}\}_{i\in N}} \sum_{i=1}^{N} h_i o_i u_i \\ \text{s.t.} \quad & \sum_{i=1}^{N} h_i o_i w_i \leq B \\ & \sum_{i=1}^{N} h_i o_i \leq H \\ & o_i \in \{0,1\} \; \forall i = 1,..,N \end{aligned}$$

where $o_i = 1$ if the district makes an offer to teacher i, h_i is an indicator of whether teacher i accepts district j's offer, if one is made, u_i is the utility from hiring teacher i, w_i is the wage, B is the budget constraint, and H is the maximum number of teachers that can be hired.

This is a linear programming problem which can be seen as a two-constraints version of the 0-1 knapsack problem. I solve it using the algorithm proposed by Martello and Toth (2003), based on the continuous relaxation of the original problem. The solution proceeds in the following steps:

1. I write the continuous relaxation (CR) of the problem, i.e. I substitute the third constraint with the milder $0 \leq o_i \leq 1 \; \forall i$. This allows me to assign a Lagrange multiplier λ to the second constraint.

2. I re-write the CR problem as a function of (λ) on the second constraint (CR(λ)):

$$\begin{aligned} & \max_{\{\mathbf{o}\}_{i\in N}} \sum_{i=1}^{N} h_i o_i (u_i - \lambda) + \lambda H \\ \text{s.t.} \quad & \sum_{i=1}^{N} h_i o_i w_i \leq B \\ & 0 \leq o_i \leq 1 \; \forall i = 1,..,N \end{aligned}$$

   For each value of λ, this continuous relaxation is a one-constraint version of the unbounded knapsack problem (Dantzig, 1957), which can be solved using linear programming techniques. As in all linear programming problems, existence of a solution follows from the feasibility of the problem, i.e. the existence of a set of offers that satisfies both constraints, which is easily verifiable in this case.

   (a) I define a teacher's *relative payoff* as the payoff the district obtains from hiring her (net of the shadow price of relaxing the capacity constraint) per dollar of salary: $(u_i - \lambda)/w_i$. Intuitively, the relative payoff measures the "efficiency" of the hire from the standpoint of the district.

   (b) I then sort teachers in descending order of relative payoff, so that $(u_i - \lambda)/w_i \geq (u_{i+1} - \lambda)/w_{i+1}$. This ranking incorporates the fact that the payoff from a hire contains both a monetary cost, captured by the salary that must be paid, and a utility cost, which stems from the the capacity constraint becoming tighter.

   (c) I define the "critical" teacher as the one indexed by $s(\lambda) = \min\{i : \sum_{j=1}^{i} w_j h_j > B$ or $h_i(u_i - \lambda) < 0\}$. In words, the critical teacher is the first teacher whose hire is unworthy from the point of view of the district, either because it leads to a violation of the budget constraint, or because the payoff from the hire is smaller than the utility cost from tightening the capacity constraint. The position of the critical teacher in the ranking separates teachers whose hire is worthy ($i < s(\lambda)$) from teachers whose hire is unworthy ($i > s(\lambda)$).

3. Solve the CR(λ) problem as in Dantzig (1957). The solution to this problem, o^{c^*}, is as follows:

$$o^{c^*}(\lambda) = \begin{cases} 1 & \text{if } i < s(\lambda) \\ \frac{c-\sum_{j=1}^{i-1} w_j}{w_i} & \text{if } i = s(\lambda) \text{ and } h_i(u_i - \lambda) \geq 0 \\ 0 \text{ if } i = s(\lambda) & \text{and } h_i(u_i - \lambda) < 0 \text{ or } i > s(\lambda) \end{cases} \tag{43}$$

where teachers are sorted so that $(u_i - \lambda)/w_i \geq (u_{i+1} - \lambda)/w_{i+1}$, and $s(\lambda)$ represents the "critical" teacher, i.e. $s(\lambda = \min\{i : \sum_{j=1}^{i} w_j h_j > B$ or $h_i(u_i - \lambda) \leq 0\})$.

4. For a given λ, define the solution to the discrete choice model as $o^*(\lambda) = \lfloor o^{c^*}(\lambda) \rfloor$.[65]

5. Select the optimal λ^* as follows:

   (a) Construct a set S of admissible levels of λ. As shown by Martello and Toth (2003), this includes: a) $\lambda = 0$; b) $\lambda = u_i \; \forall i$; c) $\lambda = (u_j w_i - u_i w_j)/(w_j - w_i) \; \forall i < j$ and such that $\lambda > 0$.

   (b) For each of these admissible levels, compute the value of the relaxed capacity constraint: $R(\lambda) = \sum_{i=1}^{s(\lambda)-1} h_i + o^{c^*}_{s(\lambda)}(\lambda) h_{s(\lambda)}$

   (c) Select the median value of the elements of S, called λ^M.

   (d) If $R(\lambda) = H$, then $\lambda^* = \lambda^m$.

   (e) If $R(\lambda) > H$, then remove λ^m from S, and reiterate from (c).

   The solution to the problem is then given by $\mathbf{o}^*(\lambda^*)$. It should be noted that this procedure selects only one value of λ^*, and therefore yields a unique solution to the problem. In principle all λ satisfying $R(\lambda) = H$ are optimal, which would give rise to multiple optimal solutions to the problem. Since, as shown by Martello and Toth (2003), $R(\lambda)$ is a non-increasing function of λ, the only case in which this can happen is if the function $R(\lambda)$ is flat and equal to H over an interval $[\underline{\lambda}, \overline{\lambda}]$. In this case, all λ in this interval would give rise to optimal solutions. It should be noted, however, that these λ will all be relatively close to each other, and virtually yield the same solution to the problem.

Estimation Procedure

I estimate the parameter vectors α (teachers' utility), α_0 (teachers' outside option), β (districts' payoff), and σ^2 (variance of the district's shock) using a maximum likelihood estimator. To do so I use data from 410 Wisconsin districts for the year 2014. For computational simplicity, I divide Wisconsin in twelve separate geographic labor markets, corresponding to the twelve Cooperative Educational Service Agencies (CESAs).[66] I consider each market as a separate one: Teachers can only move within CESAs, and

[65] This solution method corresponds to the graphical solution to the knapsack problem provided by Dantzig (1957). All teachers are represented as points in a plane having utility net of λ on the vertical dimension and salary on the horizontal dimension. Each teacher has coordinates $(u_i - \lambda, w_i)$. The solution consists in rotating clockwise a ray with the origin as pivot point and the vertical axis as starting point, and in setting $o_i^*(\lambda) = 0$ for all teachers swept out by the ray, until the point in which the sum of their salaries exceeds the budget.

[66] CESAs providing services to Wisconsin schools, such as professional development for teachers, support staff, principals, and district leaders; cooperative purchasing to maximize available resources; sharing of special professionals to meet student needs. Each CESA works with schools and districts in a particular region of Wisconsin. CESAs have no taxing authority and they receive no state aid, and rely on the fees from service contracts with member districts, and from federal, state or other grants.

districts can only make offers to teachers already working in their CESA. In 2014, about 60 percent of movements of teachers happened within a CESA.

The estimation procedure develops in the following steps (t denotes year).

1. I draw the idiosyncratic shocks ε_{ijt} for each teacher, district, and year, from a standard normal distribution. I set the initial values of the parameters to α^0, α_0^0, β^0, and σ^0.
2. For each district, teacher, and year I compute the district's payoff as $u_{ijt} = \beta^0 x_{it} + \sigma^0 \varepsilon_{ijt}^z$.
3. For each district and teacher, I compute the probability that the offer is accepted (conditional on an offer being made), h_{ijt}, with an iterative procedure:
   (a) I initially set h_{ijt} as being the probability that teacher i accepts the offer of district j when *all* districts make her an offer. The distributional assumption on ξ_{ijt} allows to write this probability as $h_{ijt}^{start} = \exp\{\alpha z_{ijt} - \alpha_0\}/(1 + \sum_{k\in J} \exp\{\alpha z_{ijt} - \alpha_0\})$.
   (b) Given h_{ijt}^{start}, I solve each district's problem and derive their optimal set of offers, o_{jt}^*, and the consequent set of offers received by each teacher, O_{it}.
   (c) Given O_{it}, I re-estimate the probability that an offer is accepted, if one is made, as $h_{ijt} = \exp\{\alpha z_{ijt} - \alpha_0\}/(1 + \sum_{k\in O_{it}} \exp\{\alpha z_{ijt} - \alpha_0\})$.
   (d) I iterate the process until convergence.
4. Having determined u_{ijt}, h_{ijt}, and w_{ijt}, I solve each district's problem. In this way I obtain the optimal set of offers, o_{jt}^*, and the consequent set of offers received by each teacher, O_{it}.
5. I write the log-likelihood function as

$$l(\alpha, \alpha_0, \beta, \sigma^2|\mathbf{z}, \mathbf{x}) = \sum_{i=1}^{N} \log h_{ij(it)t} = \sum_{i=1}^{N} \log \left[\frac{\exp\{\alpha z_{ijt} - \alpha_0\}}{1 + \sum_{k\in O_i} \exp\{\alpha z_{ijt} - \alpha_0\}} \right] \tag{44}$$

   where $j(it)$ denotes the district to which teacher i is matched in year t.

I estimate the parameters α, α_0, β, and σ using maximum likelihood. The log-likelihood is maximized using a Nelder-Mead simplex algorithm as described in Lagarias et al. (1998). I compute standard errors as the square root of the diagonal elements of the inverse of the information matrix, and I calculate the information matrix using numerical derivatives.[67]

Model Fit

To assess the within-sample fit of the model I compare a set of moments obtained simulating the model on data from 2014, used in estimation, with the same set of moments taken from the data. I use the share of teachers moving to a different district and the share of teachers exiting public schools, for all teachers and separately for teachers with positive and negative value-added. Columns 1 and 2 of Table D1 show the results of this exercise. The model tends to underestimate movements and exit. To assess out-of-sample fit I simulate the same moments using data from 2010 (not used in estimation), and compare them with moments from the data (Todd and Wolpin, 2006). Despite underpredicting movements and exits overall, the model predicts a higher moving and exit rate in 2010 compared to 2014, a pattern that is observed in the data.

[67] Optimization is implemented using the package *fminsearch* in Matlab.

Table D1: Model fit

moment	2014 (estimation year) model (1)	2014 (estimation year) data (2)	2010 (testing year) model (3)	2010 (testing year) data (4)
p(move)	0.0046	0.0145	0.0012	0.0037
VA>0	0.0062	0.0164	0.0014	0.0037
VA<0	0.0030	0.0127	0.0009	0.0037
p(exit)	0.0113	0.0532	0.0058	0.0339
VA>0	0.0113	0.0468	0.0060	0.0333
VA<0	0.0113	0.0594	0.0057	0.0344

Notes: Estimation year refers to 2014; testing year refers to 2010. Model estimates are obtained from the model and using the parameter estimates in Table 8.