Supplemental Appendix for Attracting and Retaining Highly Effective Educators in Hard-to-Staff Schools by Andrew Morgan, Minh Nguyen, Eric Hanushek, Ben Ost, and Steven Rivkin

Appendix A: PEI and TEI institutional background

The district introduced the Principal Excellence Initiative (PEI) during the 2012-2013 academic year and the Teacher Excellence Initiative (TEI) during the 2014-2015 academic year. Though they differ in many details, the two reforms share a similar structure. Each contains an achievement component based on standardized assessments, a performance component based largely on supervisor observations and judgements, and a survey component based on feedback from students or families. PEI and TEI delineate in great detail the requirements of the initiatives, points awarded for each criterium, and educator responsibilities for carrying them out. There are target distributions for ratings categories and the components of TEI and PEI to limit evaluation inflation and retain control over the personnel budget.

We now highlight some main features of TEI and PEI along with relevant implementation details. The PEI evaluation component is determined by both overall achievement and success at reducing the achievement gap. The district developed numerous assessments to measure achievement in subjects and grades lacking a state-standardized test. Initially three separate achievement scores were calculated, and the number of points assigned was the highest from three alternatives: Status (percentage of tests with scores at a specified standard); a value-added measure; and achievement score relative to the scores of a designated peer group of schools based on prior achievement. Subsequently, the status alternative was capped, and the higher point values had to be based on the value-added or peer group measures. The number of achievement points also depends on success at reducing achievement gaps by race and ethnicity. This codifies the objective of equity and support for students in demographic groups that have lower average achievement in the district and state.

PEI places substantial weight on whether a principal is an effective instructional leader. Almost 20 percent of the performance component focuses directly on improving teacher effectiveness and congruence between teacher performance and student achievement. Thus, the principal is rated on their work in support of teachers and the alignment between the subjective teacher evaluation and teacher effectiveness at raising achievement. The congruence component of the evaluation is designed to mitigate the tendency to inflate more subjective evaluations and to deter arbitrary judgements of teachers based on factors other than the quality of teaching.[1] Unlike the case for TEI, attendance and enrollment also contribute to the performance score for principals.

TEI has a similar structure as PEI, but naturally there are important differences between teacher and principal evaluation systems. Student surveys, student achievement and supervisor evaluations combine to determine the evaluation score and rating, though each of the latter two components may not count for teachers who either do not teach in tested subjects or grades or

---

[1] Morgan (2020) investigates evaluation inflation including the impact of the congruence component.

who work with students below the grade level at which surveys are administered. Supervisor classroom observations constitute the primary source of evidence for the performance score. TEI specifies ten, 10- to 15-minute spot observations of each teacher and one 45-minute extended observation per year by the designated supervisor, typically the principal or assistant principal. The supervisor is required to provide written feedback following all observations and conference with the teacher following the extended observation. Most students in grades 3-12 complete two surveys, one online and one in paper. Results from the surveys will be summarized by a statistic for teachers with sufficient number of responses. Points are assigned based on the target distribution at grade-level to assure equity because early grade-level students tend to provide more positive responses.

The achievement score is based on the results for a teacher's students (when available) and the outcomes for the entire school.[2] This is intended to foster collaboration and a common mission, but it likely also handicaps teachers who work in schools with a high fraction of ineffective educators. This may exacerbate difficulties of attracting and retaining teachers in low-performing schools, the problem ACE was designed to remedy.

Though each rating is assigned a salary, there are other considerations that can override this process. First, experience and education determine the salary for teachers new to Dallas ISD; second, teachers with fewer than three years of experience are limited in the maximum rating and compensation they can receive; and third, teachers who taught in Dallas ISD prior to the TEI reform cannot have their nominal pay lowered below its pre-reform level.

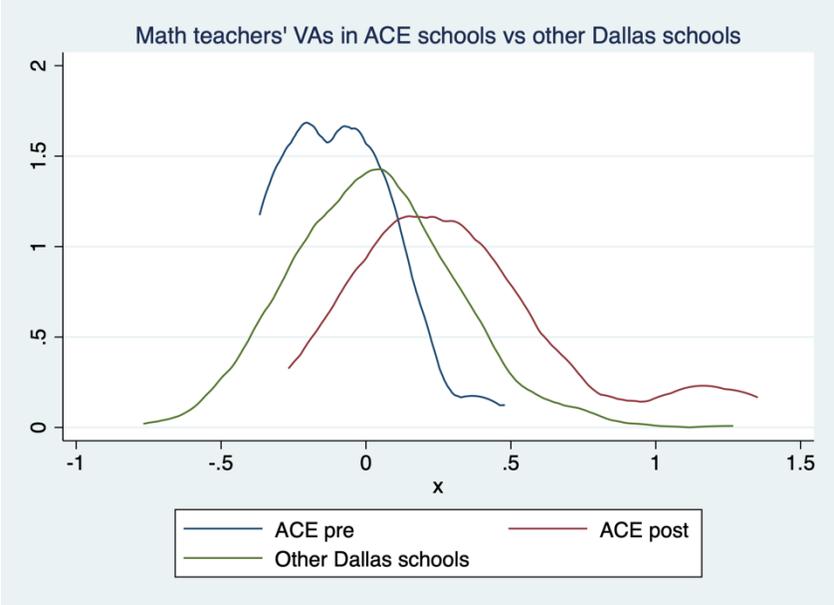[2] Since we focus on mathematics and reading/language arts teachers in tested grades, their evaluations include an achievement component based on classroom achievement.

Figure A1

A: Math

Math teachers' VAs in ACE schools vs other Dallas schools

ACE pre
ACE post
Other Dallas schools

B: Reading

Reading teachers' VA in ACE schools vs other Dallas schools

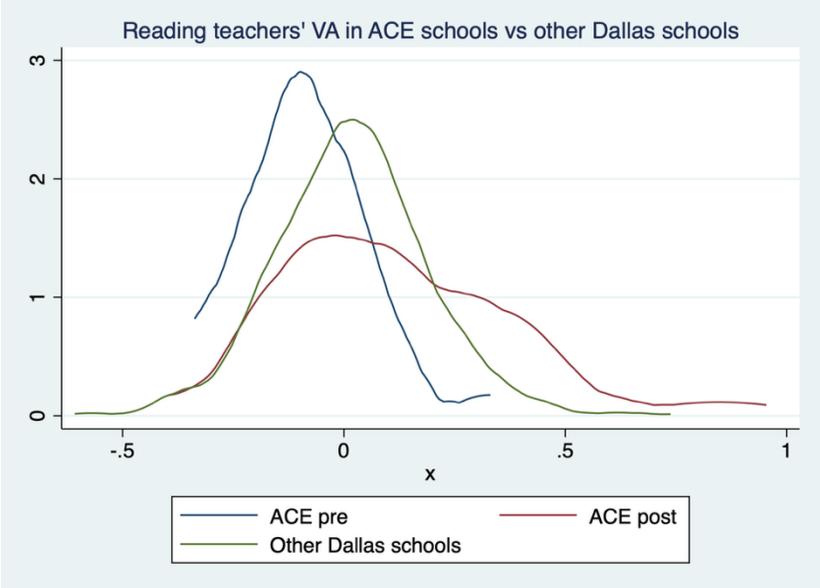ACE pre
ACE post
Other Dallas schools

Figure A2. Kernel Density Plots of 6th grade math and reading achievement for ACE 1 and control students who attended an ACE or Control school for three successive years for the final pre-treatment cohort and the fully treated cohort

A. Math

2016   2019

B. Reading

2016   2019
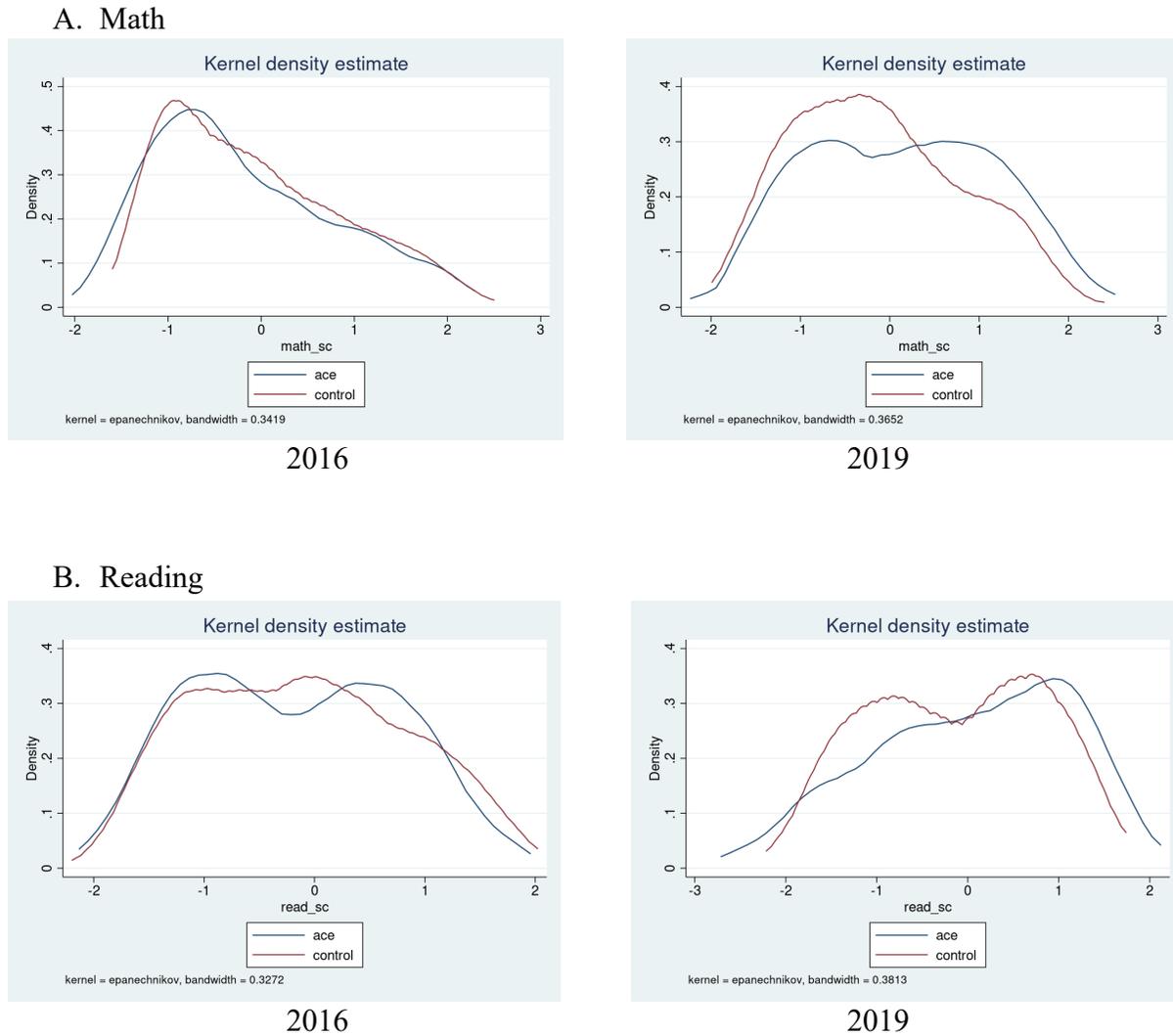
Notes: The figures present the distribution of 6th grade math and reading achievement. The 2016 figures correspond to 6th grade outcomes for students who attend an ACE or control school for 3 years prior to the adoption of ACE 1 and so they are fully untreated. The 2019 figures correspond to the 6th grade outcomes of students who attending an ACE 1 elementary school for 2016, 2017 and 2018.

Table A1: Compensation and target distribution tied with teacher effectiveness levels in the initial

year of TEI

	Unsatisfactory	Progressing		Proficient			Exemplary		Master
		I	II	I	II	III	I	II	
Salary at non-ACE school	$45K	$49K	$51K	$54K	$59K	$65K	$74K	$82K	$90K
Stipend at ACE	$0	$6K	$6K	$8K	$10K	$10K	$10K	$10K	$10K
Salary at ACE school	$45K	$55K	$57K	$62K	$69K	$75K	$85K	$92K	$100K

Source: Teacher Guidebook p36.

Table A2. Correlations between VA estimates and evaluation scores for all teachers, by subject and year

	Total evaluation score		Performance points	
	2015-16	2017-18	2015-16	2017-18
Math VA	0.616	0.552	0.363	0.285
Reading VA	0.446	0.391	0.244	0.171

Notes: The table shows the correlation between math and reading value added and the Dallas evaluation system score and performance points. All results are weighted by the number of test takers.

Table A3: Robustness to various fixed effects

Specification:	Baseline	School FE	Schl-grd FE	Student FE
ACE 1 immediate effect on math	0.5	0.524	0.522	0.516
	[0.000]	[0.000]	[0.000]	[0.002]
Observations	45,014	45,014	45,014	45,014
ACE 1 immediate effect on reading	0.255	0.267	0.275	0.284
	[0.012]	[0.004]	[0.004]	[0.004]
Observations	45,014	45,014	45,014	45,014
ACE 2 immediate effect on math	0.426	0.427	0.425	0.368
	[0.010]	[0.006]	[0.008]	[0.004]
Observations	48,957	48,957	48,957	48,957
ACE 2 immediate effect on reading	0.192	0.210	0.210	0.142
	[0.028]	[0.024]	[0.020]	[0.028]
Observations	48,950	48,950	48,950	48,950

Notes: The table shows robustness of the estimated ACE effect to including various levels of fixed effects.
Permutation test p-values shown in brackets.

Table A4: "Effects" of ACE 1 on student characteristics and predicted test scores

	Asian	Black	Hispanic	Low-income	LEP	Special ed	Male	Predicted Math	Predicted Reading
2012	-0.004	-0.039	0.034	-0.011	0.014	0.026	-0.032	0.007	0.012
	[0.526]	[0.436]	[0.432]	[0.666]	[0.742]	[0.224]	[0.274]	[0.788]	[0.544]
2013	0.001	0.016	-0.018	0.001	-0.034	0.026	-0.026	-0.018	-0.009
	[0.800]	[0.632]	[0.562]	[0.964]	[0.206]	[0.156]	[0.368]	[0.310]	[0.588]
2014	0.002	0.030	-0.027	-0.004	-0.027	0.016	-0.018	-0.019	-0.011
	[0.718]	[0.224]	[0.294]	[0.846]	[0.062]	[0.292]	[0.318]	[0.144]	[0.366]
2015	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.
2016	-0.005	0.034	-0.016	-0.023	-0.007	-0.003	0.007	-0.013	-0.007
	[0.276]	[0.150]	[0.464]	[0.418]	[0.726]	[0.854]	[0.736]	[0.324]	[0.564]
2017	-0.005	-0.043	0.055	-0.009	0.051	0.008	-0.016	0.019	0.017
	[0.418]	[0.164]	[0.124]	[0.822]	[0.132]	[0.636]	[0.594]	[0.234]	[0.246]
2018	-0.004	-0.073	0.090	-0.036	0.045	0.009	-0.040	0.034	0.038
	[0.598]	[0.238]	[0.170]	[0.216]	[0.458]	[0.640]	[0.166]	[0.238]	[0.128]
2019	0.001	-0.011	0.009	-0.021	-0.020	0.015	-0.003	-0.002	0.002
	[0.836]	[0.890]	[0.888]	[0.406]	[0.780]	[0.588]	[0.906]	[0.956]	[0.916]
N	46,933	46,933	46,933	46,933	46,933	46,933	46,933	46933	46933

Notes: The table presents event study estimates testing for compositional change at ACE 1 relative to a control group. 2015 is the year before treatment and is the reference year. Predicted math and predicted reading are generated as described in the text. Permutation test p-values shown in brackets.

Table A5: "Effects" of ACE 2 on student characteristics and predicted test scores

	Asian	Black	Hispanic	Low-income	LEP	Special ed	Male	Predicted Math	Predicted Reading
2012	-0.002	-0.037	0.044	0.048	0.021	-0.003	0.026	0.018	0.009
	[0.834]	[0.614]	[0.530]	[0.144]	[0.688]	[0.936]	[0.372]	[0.348]	[0.656]
2013	-0.004	-0.011	0.009	0.041	-0.005	0.003	-0.016	0.010	0.009
	[0.566]	[0.854]	[0.866]	[0.154]	[0.914]	[0.878]	[0.634]	[0.532]	[0.586]
2014	0.000	-0.065	0.051	0.023	0.030	0.006	0.029	0.019	0.013
	[0.910]	[0.162]	[0.348]	[0.402]	[0.454]	[0.742]	[0.276]	[0.196]	[0.368]
2015	0.001	-0.017	0.004	0.029	0.007	-0.019	0.025	0.010	0.004
	[0.710]	[0.554]	[0.900]	[0.304]	[0.834]	[0.254]	[0.388]	[0.242]	[0.638]
2016	-0.001	-0.003	-0.009	-0.003	0.002	-0.007	0.006	0.002	0.003
	[0.748]	[0.890]	[0.764]	[0.928]	[0.938]	[0.618]	[0.790]	[0.760]	[0.658]
2017	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.
2018	0.006	0.034	-0.045	0.034	-0.028	-0.009	-0.021	-0.003	-0.002
	[0.180]	[0.370]	[0.250]	[0.152]	[0.412]	[0.612]	[0.254]	[0.808]	[0.838]
2019	0.007	-0.012	-0.009	0.052	0.002	-0.004	0.002	0.009	0.004
	[0.236]	[0.724]	[0.764]	[0.014]	[0.956]	[0.876]	[0.906]	[0.450]	[0.752]
N	50,996	50,996	50,996	50,996	50,996	50,996	50,996	50,996	50,996

Notes: The table presents event study estimates testing for compositional change at ACE 2 relative to a control group. 2017 is the year before treatment and is the reference year. Predicted math and predicted reading are generated as described in the text. Permutation test p-values shown in brackets.

Table A6: Robustness to variations on definition of treatment and control

	Alternative control groups					Alternative treatment groups	
	15% (preferred)	5%	10%	20%	25%	ACE 1 includes Umphrey Lee	ACE 2 includes Onesimo and J. W. Ray
ACE 1 immediate effect on math	0.500	0.477	0.499	0.452	0.446	0.518	
	[0.012]	[0.030]	[0.008]	[0.004]	[0.000]	[0.000]	
Observations	45,019	19,545	33,064	60,359	73,590	46,554	
Number of schools	25	11	18	33	40	26	
ACE 1 immediate effect on reading	0.255	0.225	0.247	0.245	0.248	0.288	
	[0.022]	[0.152]	[0.034]	[0.026]	[0.008]	[0.002]	
Observations	45,014	19,542	33,060	60,358	73,590	46,553	
Number of schools	25	11	18	33	40	26	
ACE 2 immediate effect on math	0.426	0.473	0.442	0.402	0.396		0.452
	[0.012]	[0.030]	[0.018]	[0.010]	[0.012]		[0.000]
Observations	48,957	23,946	35,711	63,119	75,661		50,971
Number of schools	25	11	18	33	40		27
ACE 2 immediate effect on reading	0.192	0.216	0.190	0.177	0.168		0.209
	[0.044]	[0.068]	[0.066]	[0.050]	[0.048]		[0.000]
Observations	48,950	23,943	35,701	63,122	75,654		50,975
Number of schools	25	11	18	33	40		27

Notes: The table shows robustness of the estimated ACE effect to alternative control and treatment groups. The preferred specification is shown in column 1. In that specification, the control group is based on being in the lowest 15% of schools in t-2 and the treatment groups exclude Umphrey Lee from ACE 1 and exclude Onesimo Hernandez Elementary and J. W. Ray Learning Center from ACE 2. Columns 2 through 5 show how estimates change if the control group is instead based on the lowest 5%, 10%, 20%, and 25%, and columns 6 and 7 show estimates when Umphrey Lee or Onesimo Hernandez Elementary and J. W. Ray Learning Center are included in the analysis.

Table A7: Falsification test for "effect" of false 2014 treatment

	Math	Reading
2012	0.067	-0.069
	[0.308]	[0.190]
2013	ref	ref
2014	0.089	0.016
	[0.284]	[0.820]
2015	0.032	-0.126
	[0.818]	[0.088]
Observations	24,879	24,876

Notes: The table shows estimates analogous to Table 4, but considers the effect of a false treatment set to occur in 2014 and assigned based on 2012 performance. Permutation test p-values shown in brackets.

Table A8: Robustnes to excluding schools that lose teachers to ACE

	Preferred	Excluding control schools that lose a teacher to ACE
ACE 1 immediate effect on math	0.5	0.532
	[0.000]	[0.018]
Observations	45,019	26,005
ACE 1 immediate effect on reading	0.255	0.282
	[0.012]	[0.048]
Observations	45,014	26,000
ACE 2 immediate effect on math	0.426	0.489
	[0.010]	[0.002]
Observations	48,957	26,709
ACE 2 immediate effect on reading	0.192	0.199
	[0.028]	[0.048]
Observations	48,950	26,700

Note: Column 1 shows the preferred approach and the column 2 shows how the estimates change when excluding schools that lose a teacher to ACE from the control group. Permutation test p-values are shown in brackets.

Table A9: Inference using cluster wild-bootstrap

	ACE 1		ACE 2	
	Math	Reading	Math	Reading
2012	0.036	0.128	-0.071	-0.035
	[0.510]	[0.240]	[0.398]	[0.616]
2013	0.055	-0.040	-0.171	-0.024
	[0.712]	[0.852]	[0.078]	[0.598]
2014	0.012	-0.002	-0.034	0.032
	[0.922]	[0.988]	[0.688]	[0.748]
2015	ref.	ref.	0.093	0.053
			[0.136]	[0.764]
2016	0.500	0.255	-0.053	-0.031
	[0.014]	[0.052]	[0.652]	[0.774]
2017	0.580	0.352	ref.	ref.
	[0.028]	[0.052]		
2018	0.478	0.353	0.426	0.192
	[0.062]	[0.086]	[0.018]	[0.112]
2019	0.280	0.186	0.497	0.320
	[0.506]	[0.318]	[0.038]	[0.042]
Observations	45,019	45,014	48,957	48,950

Notes: P-values based on the cluster wild bootstrap approach shown in brackets. For ACE 1, the program is no longer in effect starting in 2019.