Revisiting the Impacts of Teachers
Online Appendix

Jesse Rothstein[*]

January 25, 2017

A Reproduction of CFR-I Results

Appendix Tables A1-A7 present CFR-I's (CFR 2014*a*) results from New York in parallel with reproductions, using CFR's (2014*c*) code, in data from North Carolina.

Table A1 presents student-level summary statistics (from CFR-I's Table 1, Panel A). Free lunch and minority shares are lower in North Carolina than in New York, but (surprisingly) the recorded English language learner share is higher. In North Carolina, this variable and special education status are missing from 2009 onward; summary statistics pertain only to those with

[*] Goldman School of Public Policy and Department of Economics, University of California, Berkeley, rothstein@berkeley.edu. Additional analyses, supplementary materials, and replication files are available at `http://eml.berkeley.edu/~jrothst/CFR/`

non-missing data.

Table A2 presents CFR-I's Table 2. Autocovariances are similar in the two samples for elementary English teachers, but higher in the North Carolina sample for elementary math teachers. Similarly, in English the two samples yield nearly identical estimates of the standard deviation of teachers' VA, net of sampling error, but in math the North Carolina sample yields an estimate about one-fifth larger than does CFR-I's sample.

Figure A1 displays the autocorrelations graphically. In both samples, the autocorrelations are higher in math than in reading; they are also higher in each subject in North Carolina than in CFR-I's sample. Where CFR-I found that the autocorrelations stabilize at lags longer than 7, the North Carolina sample suggests that they continue to decline out to the end of the sample.

Table A3 presents results from CFR-I's Table 3. (I do not reproduce their Column 3, as their code archive does not make clear how their dependent variable is constructed.) Results are broadly similar. In Column 2, my coefficient (0.009) is significantly different from zero where theirs (0.002) is not, but both are small in magnitude. Table A4 presents estimates from CFR-I's Table 4. Many of these are presented elsewhere as well; they are included here for completeness. I do not reproduce CFR-I's Column 5, as my North Carolina sample excludes middle school grades. Again, all estimates are strikingly similar between the two samples. Table A5 presents estimates from CFR-I's Table 5. Estimates are quite similar, despite the higher share of teachers assigned predicted VA scores of zero in Column 2 in my sample (27.4%) than in CFR-I's (16.4%). Rothstein (2017) presents additional relevant results.

Table A6 reproduces CFR-I's Table 6. Notably, the North Carolina results indicate *negative* forecast bias in rows 1-6. But results are generally quite similar.

Finally, Table A7 presents selected estimates from Table 2 in CFR-I's online appendix. These are coefficients of regressions of student characteristics on their teachers' predicted VA. Raw regression coefficients are attenuated because the predicted VA measures are shrunken, and thus have lower variance than the teachers' true effects. CFR-I multiply their coefficients by 1.56, the average ratio of the standard deviation of true effects to the standard deviation of predicted effects. In North Carolina, this ratio is 1.36, so coefficients in Panel B are multiplied by this. Estimates are broadly similar, though there is perhaps less sorting of high-prior-achievement students to high-predicted-VA teachers in North Carolina than in CFR-I's sample. One notable difference is that minority students have lower-predicted-VA teachers, on average, than non-minority students in North Carolina, but not in New York.

B Additional specifications

Responding to an early draft of this comment, CFR (2014*b*) suggested that the failure of the placebo test might be due to so-called "mechanical" effects

– to factors that influence both prior year scores and measured teacher VA (but perhaps not actual teacher effectiveness). Specifically, CFR note that data from $t-2$ is used both to predict the VA of teachers in $t-1$ and t, and thus to compute ΔQ_{sgmt}, and for the prior-year scores of $t-1$ students. This could create a spurious correlation between ΔQ_{sgmt} and the change in prior year scores. In Table 2 I found that the placebo test failed even when only non-test outcomes were used to measure student preparedness. This demonstrates that test dynamics cannot possibly account for the result. Nevertheless, in Table A8 I explore several alternative specifications aimed at removing the specific mechanical effects that CFR suggest.

Row 1 presents baseline estimates, repeated from Tables 2 and 3. Row 2 is identical but with standard errors clustered at the school level; this increases standard errors by about one-third.[1]

CFR (2014*b*; 2015) suggest that one source of potential mechanical effects is teachers who teach the same cohort of students in multiple years as they progress across grades. If a teacher taught in grade $g-1$ in $t-2$ and then taught the same students in grade g in $t-1$, then the both the average VA in grade g in $t-1$ (and thus ΔQ_{sgmt}) and the average lagged scores of grade g students in $t-1$ will reflect her effectiveness.[2] CFR (2014*b*)

[1] CFR-I's main results cluster at the school-by-cohort level. School-level clustering is more general. Moreover, I present below IV specifications with school-year fixed effects; it is computationally difficult to cluster these at the school-cohort level.

[2] This is a source of a mechanical association in the differenced specification only if the teacher leaves the school or grade in t; otherwise, her VA does not contribute to the $t-1$ to t change. Note also that "following" is a problem for the quasi-experimental analysis as well as for the placebo test. The quasi-experimental analysis is designed to test whether

propose addressing this by instrumenting for the change in VA, ΔQ_{sgmt}, with a modified measure that excludes teachers who taught $g-1$ in $t-2$ or $t-1$. This is implemented by setting predicted VA for these teachers to zero.

In North Carolina, less than 4% of teacher mobility consists of teachers following students. Not surprisingly, when I modify ΔQ_{sgmt} to exclude teachers who taught grade $g-1$ in $t-2$ or $t-1$, or who taught grade $g-2$ in $t-3$ or $t-2$, the modification makes little difference. The modified version of ΔQ_{sgmt} is correlated 0.96 with the original version, and the first-stage coefficient is 0.98. Estimates of my key specifications are shown in Row 3 of Table A8. When classrooms with missing VA scores are excluded, the association with the change in prior-year scores is reduced but remains significant, and the λ estimate is hardly changed. Note that the no-follower instrument involves setting some teachers' VA predictions to the grand mean, and thus relies on the same assumption of within-school independence as does the inclusion of teachers with missing leave-two-out predictions, also set to the grand mean. There is thus no set of assumptions that can justify the subsample specifications in columns 1-3. When all classrooms are included, in columns 4-6, the placebo test coefficient is no longer significant, but the λ coefficient from a specification without controls falls to match that in the specification with

VA scores accurately forecast the impact of grade-g teachers on their students' learning in grade-g; if a portion of the $\hat{\lambda}$ coefficient reflects contributions that the same teachers made to students when they were in grade $g-1$, this would need to be controlled in order to isolate the causal effect of interest.

controls. I thus conclude that "follower" teachers might contribute slightly to the placebo test violation, but that recognition of this phenomenon has no effect on my conclusions regarding forecast bias.[3]

CFR (2014*b*; 2015) also suggest that school-year-subject shocks could create mechanical, spurious failures of the placebo test: A positive shock to a school in $t-2$ will raise both the predicted VA of the school's $t-1$ teachers and the prior-year scores of the $t-1$ students. This would be absorbed by school-year effects already included in the main specifications if it were common across subjects, but subject-specific shocks would not be. CFR (2014*b*; 2015) propose to address it by including school-subject-year fixed effects. I implement this in Row 4. This halves the number of degrees of freedom, leaving only three or fewer observations per cell. Standard errors are larger here. The quasi-experimental estimates in Columns 2 and 3 rise, and I cannot reject $\lambda = 1$ in Column 3. However, in the preferred sample that includes all classrooms (assigning VA predictions of zero to teachers with missing data), the additional fixed effects make little difference at all, and I decisively reject $\lambda = 1$. Row 5 presents a specification with both school-subject-year effects and instrumentation for follower teachers. The

[3] I have also explored specifications analogous to those in Columns 3 and 6 where I instrument for the change in mean prior-year scores with a modified version that excludes students of teacher "followers." This has no effect on the results. When CFR (2015) estimate the specification in Column 1, the coefficient is insignificantly different from zero, though this coefficient is significant in Los Angeles (Bacher-Hicks, Kane and Staiger, 2014). This may be the sole substantively important difference in empirical results across the three samples. In any event, when CFR (2015) use the "no followers" design for the main quasi-experimental specification (as in Column 2), they estimate $\hat{\lambda} = 0.92$ and reject the null hypothesis that $\lambda = 1$. This is quite similar to my results.

main placebo test coefficient is insignificant here, but my preferred forecast bias coefficient (in column 6) is unchanged, at 0.89, and remains significantly different from 1.

The inclusion of school-subject-year effects is not the only way to address the possibility that common shocks would affect both teachers' VA predictions and students' lagged scores. An alternative, more consistent with the overall research design, is to exclude $t-2$ data from the predictions of teacher VA in years $t-1$ and t. "Leave-three-out" VA predictions, ensure that there is zero overlap between the scores used to construct the VA scores and those used for the dependent variable in the placebo test, as the latter is based only on data from $t-2$ and $t-1$. Row 6 presents estimates using these leave-three-out VA predictions. They are quite similar to the baseline estimates, if anything indicating larger selection problems and smaller quasi-experimental estimates. Row 7 combines the leave-three-out VA scores with the no-follower IV, with quite similar results

CFR (2015) point out that with serial correlation in the school-year-subject shocks, a shock in $t-3$ would influence leave-three-out VA scores and be correlated with the shock to prior-year scores for the $t-1$ cohort, potentially biasing leave-threeee-out placebo test. Such serial correlation would create a similar bias in the CFR-I quasi-experiment, as $t-2$ shocks enter into VA scores and would be similarly correlated with the shock to $t-1$ scores, and indeed one would expect the leave-three-out strategy to reduce bias.

Nevertheless, rows 8 and 9 present estimates that use leave-four-out and leave-five-out VA scores that exclude not just $t-2$ but also $t-3$ and (in Row 9) $t-4$ data from the calculations. Results are extremely stable. In row 10, I take this to the logical extreme, using only data from $t+1$ and thereafter to forecast (backcast) VA in $t-1$ and t. This specification, proposed by CFR (2014*b*), should entirely eliminate any mechanical effect of the form that CFR (2014*b*; 2015) propose, but estimates are basically unchanged – if anything, the forecast bias coefficient falls from the baseline specification ($\hat{\lambda} = 0.83$ vs. 0.86).

Taking the various specifications in Table A8 together, along with the non-test placebo analysis in Table 2, the evidence is clear that mechanical effects cannot account for the results. Rothstein (2017) presents additional sensitivity analyses, focusing on the sample selection created by the exclusion of classrooms with missing leave-two-out teacher VA scores. Results are presented that vary the procedure for assigning VA predictions to these teachers and that limit the sample to cells with no excluded classrooms.

References

Bacher-Hicks, Andrew, Thomas J. Kane, and Douglas O. Staiger. 2014. "Validating Teacher Effect Estimates Using Changes in Teacher Assignments in Los Angeles." National Bureau of Economic Research Working paper 20657.

Chetty, Raj, John N Friedman, and Jonah E Rockoff. 2014*a*. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review*, 104(9): 2593–2632.

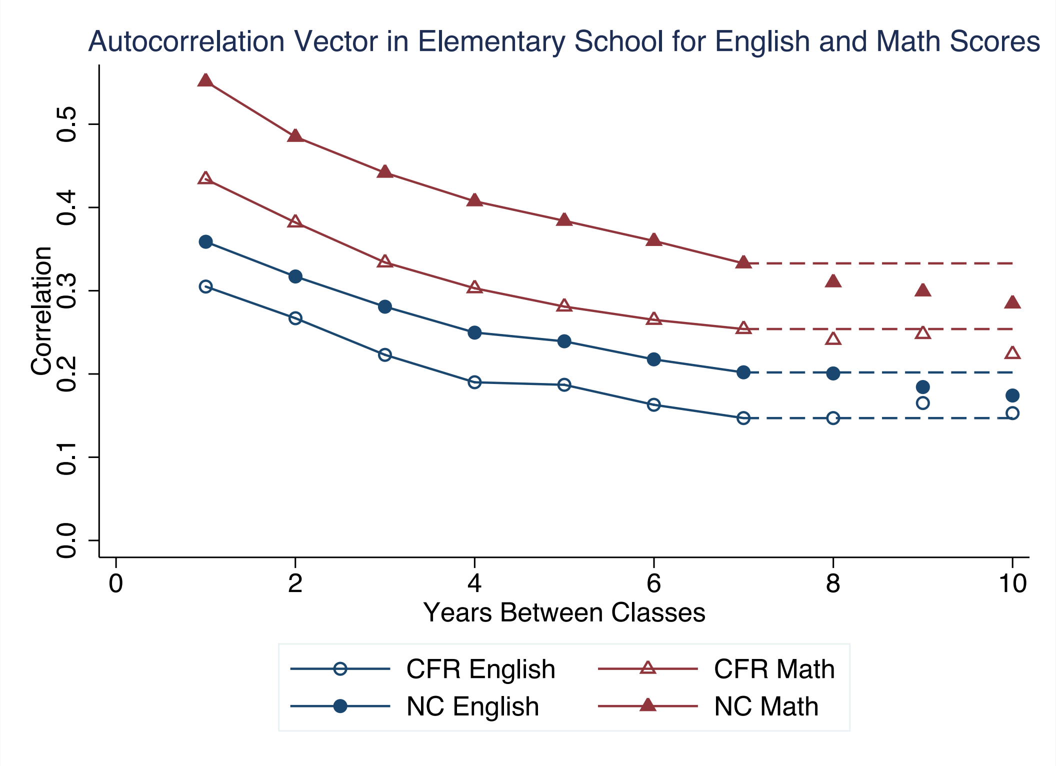
Chetty, Raj, John N Friedman, and Jonah E Rockoff. 2014*b*. "Prior Test Scores Do Not Provide Valid Placebo Tests of Teacher Switching Research Designs." Unpublished manuscript. Downloaded October 13, 2014 from http://obs.rc.fas.harvard.edu/chetty/va_prior_score.pdf.

Chetty, Raj, John N Friedman, and Jonah E Rockoff. 2014*c*. "Stata Code for Implementing Teaching-Staff Validation Technique." Downloaded July 21, 2014, from http://obs.rc.fas.harvard.edu/chetty/cfr_analysis_code.zip.

Chetty, Raj, John N Friedman, and Jonah E Rockoff. 2015. "Measuring the Impacts of Teachers: Response to Rothstein (2014)." July. Unpublished manuscript. Downloaded July 27, 2015 from http://obs.rc.fas.harvard.edu/chetty/va_response.pdf.

Rothstein, Jesse. 2017. "Supplement to "Revisiting the Impact of Teachers"." Manuscript. Available at http://eml.berkeley.edu/~jrothst/CFR/rothstein_CFR_supplement_jan2017.pdf.

Appendix Figure 1
Reproduction of CFR-I, Figure 1A



Notes: See notes to CFR-I, Figure 1.

Appendix Table A1. Reproduction of CFR-I, Table 1 (Panel A only)
Summary statistics for sample used to estimate value-added model

	CFR-I, Table 1			North Carolina sample		
	Mean	SD	N	Mean	SD	N
	(1)	(2)	(3)	(4)	(5)	(6)
Class size (not student weighted)	27.3	5.6	391,487	22.2	5.0	357,036
No. of subject-years per student	5.6	3.0	1,367,051	4.5	1.7	1,607,198
Test score (SD)	0.2	0.9	7,639,288	0.0	1.0	7,215,581
Female	50.8%		7,639,288	49.7%		7,215,581
Age (years)	11.4	1.5	7,639,288	10.5	0.9	7,213,590
Free lunch elig.	79.6%		5,021,163	44.9%		3,926,246
Minority (Black/Hispanic)	71.6%		7,639,288	34.2%		7,215,581
English language learner	4.8%		7,639,288	8.5%		5,996,113
Special education	1.9%		7,639,288	2.3%		5,478,335
Repeating grade	1.7%		7,639,288	1.4%		7,215,581
Matched to parents in tax data	87.7%		7,639,288			

Notes: See notes to CFR-I, Table 1. In New York, free lunch eligibility is available only for 1999-2009. In North Carolina, it is available only for 1999-2006, and English language learner and special education information are available only 1997-2008.

Appendix Table A2. Reproduction of CFR (2014a), Table 2
Teacher Value-Added Model Parameter Estimates

	CFR		North Carolina sample	
	Elem. School English	Elem. School Math	Elem. School English	Elem. School Math
	(1)	(2)	(3)	(4)
	Panel A: Autocovariance and Autocorrelation Vectors			
Lag 1	0.013	0.022	0.012	0.032
	(0.0003)	(0.0003)	(0.0002)	(0.0002)
	[0.305]	[0.434]	[0.359]	[0.551]
Lag 2	0.011	0.019	0.011	0.028
	(0.0003)	(0.0003)	(0.0002)	(0.0003)
	[0.267]	[0.382]	[0.317]	[0.485]
Lag 3	0.009	0.017	0.009	0.026
	(0.0003)	(0.0004)	(0.0002)	(0.0003)
	[0.223]	[0.334]	[0.281]	[0.442]
Lag 4	0.008	0.015	0.008	0.023
	(0.0004)	(0.0004)	(0.0002)	(0.0004)
	[0.190]	[0.303]	[0.250]	[0.407]
Lag 5	0.008	0.014	0.008	0.022
	(0.0004)	(0.0005)	(0.0002)	(0.0004)
	[0.187]	[0.281]	[0.239]	[0.384]
Lag 6	0.007	0.013	0.007	0.021
	(0.0004)	(0.0006)	(0.0003)	(0.0005)
	[0.163]	[0.265]	[0.218]	[0.360]
Lag 7	0.006	0.013	0.007	0.019
	(0.0005)	(0.0006)	(0.0003)	(0.0005)
	[0.147]	[0.254]	[0.202]	[0.333]
Lag 8	0.006	0.012	0.006	0.018
	(0.0006)	(0.0007)	(0.0003)	(0.0006)
	[0.147]	[0.241]	[0.201]	[0.310]
Lag 9	0.007	0.013	0.006	0.017
	(0.0007)	(0.0008)	(0.0003)	(0.0007)
	[0.165]	[0.248]	[0.184]	[0.299]
Lag 10	0.007	0.012	0.006	0.017
	(0.0008)	(0.0010)	(0.0004)	(0.0008)
	[0.153]	[0.224]	[0.174]	[0.285]
	Panel B: Within-Year Variance Components			
Total SD	0.537	0.517	0.561	0.544
Individual Level SD	0.506	0.473	0.542	0.495
Class+Teacher Level SD	0.117	0.166	0.144	0.225
Estimates of Teacher SD				
Lower Bound Based on Lag 1	0.113	0.149	0.110	0.180
Quadratic Estimate	0.124	0.163	0.118	0.192

Notes: See notes to CFR (2014a), Table 2. In Panel A, each entry includes the autocovariance, the standard error of that covariance (in parentheses), and the autocorrelation (in brackets) of average test score residuals across years, within teachers.

Appendix Table A3. Reproduction of CFR (2014a), Table 3
Estimates of Forecast Bias Using Parent Characteristics and Lagged Scores

Dep. Var.:	Score in Year t	Pred. Score using Parent Chars.	Score in Year t	Pred. Score using Year t-2 Score
	(1)	(2)	(3)	(4)
	Panel A: CFR (2014a)			
Teacher VA	0.998	0.002	0.996	0.022
	(0.0057)	(0.0003)	(0.0057)	(0.0019)
Parent Chars. Controls			X	
Observations	6,942,979	6,942,979	6,942,979	5,096,518
	Panel B: North Carolina sample			
Teacher VA	1.021	0.009		0.022
	(0.004)	(0.001)		(0.002)
Parent Chars. Controls				
Observations	5,142,680	3,584,736		3,014,172

Notes: See notes to CFR (2014a), Table 3; replication follows their methods. Dependent variables are residualized against the covariates in the VA model, at the individual level, before being regressed on on the teacher's leave-one-out predicted VA, controlling for subject. In Column 2, the second stage regression is estimated on classroom-subject-level aggregates; reported observation counts correspond to the number of student-year-subject-level observations represented in these aggregates. Standard errors are clustered at the school-cohort level.

Appendix Table A4. Reproduction of CFR (2014a), Table 4
Quasi-Experimental Estimates of Forecast Bias

Dependent Variable:	Δ Score	Δ Score	Δ Score	Δ Predicted Score	Δ Other Subj. Score	Δ Other Subj. Score
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: CFR (2014a)					
Change in mean teacher predicted VA	0.974	0.957	0.950	0.004	0.038	0.237
across cohorts	(0.033)	(0.034)	(0.023)	(0.005)	(0.083)	(0.028)
Year Fixed Effects	X				X	X
School x Year Fixed Effects		X	X	X		
Lagged Score Controls			X			
Lead and Lag Changes in Teacher VA			X			
Other-Subject Change in Mean Teacher VA					X	X
Grades	4 to 8	4 to 8	4 to 8	4 to 8	Middle Sch.	Elem. Sch.
No. of School x Grade x Subject x Year Cells	59,770	59,770	46,577	59,323	13,087	45,646
	Panel B: North Carolina sample					
Change in mean teacher predicted VA	1.097	1.030	0.994	0.008		0.202
across cohorts	(0.022)	(0.021)	(0.017)	(0.011)		(0.016)
Year Fixed Effects	X					X
School x Year Fixed Effects		X	X	X		
Lagged Score Controls			X			
Lead and Lag Changes in Teacher VA			X			
Other-Subject Change in Mean Teacher VA						X
Grades	3 to 5	3 to 5	3 to 5	3 to 5		3 to 5
No. of School x Grade x Subject x Year Cells	79,466	79,466	58,385	54,663		76,548

Notes: See notes to CFR (2014a), Table 4. Panel B replicates CFR's estimates using the North Carolina sample.

Appendix Table A5. Reproduction of CFR (2014a), Table 5
Quasi-Experimental Estimates of Forecast Bias: Robustness Checks

Specification:	Teacher Exit Only	Full Sample	<25% Imputed VA	0% Imputed VA
Dependent Variable:	Δ Score	Δ Score	Δ Score	Δ Score
	(1)	(2)	(3)	(4)
	Panel A: CFR (2014a)			
Change in mean teacher predicted VA	1.045	0.877	0.952	0.990
across cohorts	(0.107)	(0.026)	(0.032)	(0.045)
Year Fixed Effects	X	X	X	X
Number of School x Grade x Subject x Year Cells	59,770	62,209	38,958	17,859
Pct. of Observations with Non-Imputed VA	100.0	83.6	93.8	100.0
	Panel B: North Carolina sample			
Change in mean teacher predicted VA	1.174	0.936	1.100	1.081
across cohorts	(0.040)	(0.022)	(0.035)	(0.043)
Year Fixed Effects	X	X	X	X
Number of School x Grade x Subject x Year Cells	79,466	91,221	34,495	23,445
Pct. of Observations with Non-Imputed VA	100.0	72.6	94.4	100.0

Notes: See notes to CFR (2014a), Table 5. Panel B replicates CFR's estimates using the North Carolina sample.

Appendix Table A6. Reproduction of CFR (2014a), Table 6
Comparisons of Forecast Bias Across Value-Added Models

	CFR-I		North Carolina	
	Correlation with baseline VA estimates	Quasi-experimental estimate of bias (%)	Correlation with baseline VA estimates	Quasi-experimental estimate of bias (%)
	(1)	(2)	(3)	(4)
1. Baseline	1.000	2.58	1.000	-9.69
		(3.34)		(2.19)
2. Baseline, no teacher FE	0.979	2.23	0.981	-6.07
		(3.50)		(2.22)
3. Baseline, with teacher experience	0.989	6.66		
		(3.28)		
4. Prior test scores	0.962	3.82	0.976	-9.13
		(3.30)		(2.18)
5. Student's lagged scores in both subjects	0.868	4.83	0.955	-4.88
		(3.29)		(2.17)
6. Student's lagged score in same subj. only	0.787	10.25	0.923	-3.09
		(3.17)		(2.13)
7. Non-score controls	0.662	45.39	0.683	31.00
		(2.26)		(1.56)
8. No controls	0.409	65.58	0.522	46.41
		(3.73)		(1.32)

Notes: See notes to CFR-I, Table 6. CFR (2014a) do not provide code for the row 3 specification. Negative bias share coefficients in column 4 reflect estimated forecast coefficients above 1.

Appendix Table A7: Replication of CFR (2014a), Appendix Table 2
Differences in Teacher Quality Across Students and Schools

	Dependent variable: Teacher value-added						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A: CFR (2014a), Appendix Table 2						
Lagged test score	0.0122			0.0123			
	(0.0006)			(0.0006)			
Special educ. student		-0.003					
		(0.001)					
Parent income ($10,000s)			0.00084	0.00001			
			(0.00013)	(0.00011)			
Minority (black/hispanic) student					-0.001		
					(0.001)		
School mean parent income ($10,000s)						0.0016	
						(0.0007)	
School fraction minority							0.003
							(0.003)
N	6,942,979	6,942,979	6,094,498	6,094,498	6,942,979	6,942,979	6,942,979
	Panel B: North Carolina sample						
Lagged test score	0.0077						
	(0.0004)						
Special ed		0.0055					
		(0.0006)					
Minority (black/hispanic) student					-0.0028		
					(0.0012)		
School fraction minority							0.0054
							(0.0042)
N	5,142,680	5,142,680			5,142,680		5,142,680

Notes: See notes to CFR (2014a), Appendix Table 2. Panel B reports coefficients from applying CFR's code to the North Carolina sample. CFR multiply their reported coefficients by 1.56 to offset the average shrinkage of the dependent variable. The corresponding factor in the North Carolina sample (using CFR-I's calculation) is 1.36, and coefficients in Panel B are multiplied by that.

Appendix Table A8. Assessing potential mechanical contributions to the placebo test failure

		Excluding classrooms without VA predictions			Including all classrooms		
	Dependent variable	Δ Prior Year Score	Δ End-of-Year Score		Δ Prior Year Score	Δ End-of-Year Score	
			No controls	With control for Δ prior year score		No controls	With control for Δ prior year score
		(1)	(2)	(3)	(4)	(5)	(6)
1	Baseline	*0.14*	1.03	*0.93*	*0.09*	*0.90*	*0.86*
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
2	Cluster on school	*0.14*	1.03	*0.93*	*0.09*	*0.90*	*0.86*
		(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.02)
3	IV setting VA of following teachers to zero	*0.08*	1.00	*0.95*	0.03	*0.87*	*0.87*
		(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.02)
4	School-year-subject FEs	*0.12*	1.06	0.97	0.06	*0.91*	*0.89*
		(0.04)	(0.04)	(0.02)	(0.04)	(0.04)	(0.03)
5	School-year-subject FEs, IV	0.05	1.03	0.99	-0.02	*0.87*	*0.89*
		(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.02)
6	Using leave-three-out teacher VA predictions	*0.17*	1.03	*0.92*	*0.12*	*0.91*	*0.85*
		(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.02)
7	Leave-three-out, IV	*0.12*	1.01	*0.93*	*0.07*	*0.88*	*0.85*
		(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.02)
8	Using leave-four-out teacher VA predictions	*0.16*	1.02	*0.91*	*0.13*	*0.90*	*0.84*
		(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)
9	Using leave-five-out teacher VA predictions	*0.15*	1.02	*0.91*	*0.13*	*0.89*	*0.83*
		(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)
10	Using leave-past-out teacher VA predictions	*0.14*	0.99	*0.89*	*0.12*	*0.88*	*0.82*
		(0.03)	(0.04)	(0.02)	(0.04)	(0.04)	(0.03)

Notes: Specifications in Row 1 correspond to Table 2, Column 1 (Cols. 1 and 4); Table 3, Column 1 (Cols. 2 and 5); and Table 3, Column 2 (Cols. 3 and 6). In each case, Columns 1-3 correspond to the Panel A specification in the earlier table, and Columns 4-6 to the Panel B specification. Successive rows modify the specification. In Rows 2-9, standard errors are clustered at the school level. In Row 3, the change in mean predicted teacher VA in the school-grade-subject-year cell is instrumented with a variable constructed similarly but with predicted VA set to zero for teachers who have ever previously taught the same cohorts. Row 4 presents OLS estimates with school-year-subject fixed effects, while row 5 reports IV estimates of the same specification using the non-following teacher instrument. In Rows 6-9, teacher VA predictions are constructed using only data from before t-2 (rows 6 and 7), t-3 (row 8), or t-4 (row 9). In Row 10, only data from after t is used. Row 7 applies the IV specification from Row 3 to the model from row 6, using leave-3-out VA predictions for non-follower teachers. Italicized coefficients are significantly different from the null hypothesis (zero in Columns 1 and 4; one in Columns 2, 3, 5, and 6).

Appendix Table B1. Assessing potential mechanical contributions to the placebo test failure

		Excluding classrooms without VA predictions			Including all classrooms		
	Dependent variable	Δ Prior Year Score	Δ End-of-Year Score		Δ Prior Year Score	Δ End-of-Year Score	
			No controls	With control for Δ prior year score		No controls	With control for Δ prior year score
		(1)	(2)	(3)	(4)	(5)	(6)
1	Baseline	*0.14*	1.03	*0.93*	*0.09*	*0.90*	*0.86*
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
2	Cluster on school	*0.14*	1.03	*0.93*	*0.09*	*0.90*	*0.86*
		(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.02)
3	IV setting VA of following teachers to zero	*0.08*	1.00	*0.95*	0.03	*0.87*	*0.87*
		(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.02)
4	School-year-subject FEs	*0.12*	1.06	0.97	0.06	*0.91*	*0.89*
		(0.04)	(0.04)	(0.02)	(0.04)	(0.04)	(0.03)
5	School-year-subject FEs, IV	0.05	1.03	0.99	-0.02	*0.87*	*0.89*
		(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.02)
6	Using leave-three-out teacher VA predictions	*0.17*	1.03	*0.92*	*0.12*	*0.91*	*0.85*
		(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.02)
7	Leave-three-out, IV	*0.12*	1.01	*0.93*	*0.07*	*0.88*	*0.85*
		(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.02)
8	Using leave-four-out teacher VA predictions	*0.16*	1.02	*0.91*	*0.13*	*0.90*	*0.84*
		(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)
9	Using leave-five-out teacher VA predictions	*0.15*	1.02	*0.91*	*0.13*	*0.89*	*0.83*
		(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)
10	Using leave-past-out teacher VA predictions	*0.14*	0.99	*0.89*	*0.12*	*0.88*	*0.82*
		(0.03)	(0.04)	(0.02)	(0.04)	(0.04)	(0.03)

Notes: Specifications in Row 1 correspond to Table 2, Column 1 (Cols. 1 and 4); Table 3, Column 1 (Cols. 2 and 5); and Table 3, Column 2 (Cols. 3 and 6). In each case, Columns 1-3 correspond to the Panel A specification in the earlier table, and Columns 4-6 to the Panel B specification. Successive rows modify the specification. In Rows 2-9, standard errors are clustered at the school level. In Row 3, the change in mean predicted teacher VA in the school-grade-subject-year cell is instrumented with a variable constructed similarly but with predicted VA set to zero for teachers who have ever previously taught the same cohorts. Row 4 presents OLS estimates with school-year-subject fixed effects, while row 5 reports IV estimates of the same specification using the non-following teacher instrument. In Rows 6-9, teacher VA predictions are constructed using only data from before t-2 (rows 6 and 7), t-3 (row 8), or t-4 (row 9). In Row 10, only data from after t is used. Row 7 applies the IV specification from Row 3 to the model from row 6, using leave-3-out VA predictions for non-follower teachers. Italicized coefficients are significantly different from the null hypothesis (zero in Columns 1 and 4; one in Columns 2, 3, 5, and 6).

Appendix Table B2. Assessing sensitivity of results to the imputation model

	Excluding classrooms missing teacher VA predictions	Including all classrooms, assigning to teachers with missing VA predictions:			
		Grand mean	School mean	Missing mean	Missing mean at school
	(1)	(2)	(3)	(4)	(5)
	Panel A: Quasi-experimental models without controls				
Change in mean teacher predicted VA	1.030	0.904	0.915	0.933	0.911
	(0.021)	(0.022)	(0.022)	(0.022)	(0.021)
	Panel B: Models for change in prior-year scores				
Change in mean teacher predicted VA	0.144	0.092	0.134	0.084	0.128
	(0.021)	(0.022)	(0.023)	(0.023)	(0.022)
	Panel C: Models for change in end-of-year scores, with controls for change in prior-year scores				
Change in mean teacher predicted VA	0.933	0.860	0.850	0.892	0.847
	(0.015)	(0.017)	(0.017)	(0.017)	(0.017)
Change in mean student prior year score	0.675	0.536	0.535	0.536	0.535
	(0.004)	(0.009)	(0.009)	(0.009)	(0.009)

Notes: Specifications in column 1, panels A-C are identical to those in Table 1, Column 2; Table 2, Column 1; and Table 3, Column 2, respectively. Successive columns include all classrooms in the dependent and independent variables, varying the VA prediction assigned to teachers who are excluded in column 1. In column 2, these teachers are assigned the grand mean of zero. In Column 3, the prediction is based on the shrunken leave-two-out mean at the same school. In Column 4, it uses the shrunken leave-two-out mean among all teachers with missing VA predictions. In column 5, it uses the shrunken leave-two-out mean among all teachers at the school with missing VA predictions. All specifications include school-year fixed effects. N=79,466 school-grade-subject-year cells in Column 1; 91,221 in Columns 2-5 in Panel A; and 90,701 in Columns 2-5, Panels B-C.

Appendix Table B3. Robustness of CFR-I, Table 5's robustness results
Quasi-Experimental Estimates of Forecast Bias: Robustness Checks

	Teacher Exit Only		Full Sample		<25% Imputed VA		0% Imputed VA	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: Quasi-experimental models without controls							
Change in mean teacher predicted VA	1.174	1.080	0.936	0.904	1.100	0.965	1.081	0.918
	(0.040)	(0.044)	(0.022)	(0.022)	(0.035)	(0.040)	(0.043)	(0.051)
Year fixed effects	X		X		X		X	
School-year fixed effects		X		X		X		X
Number of School x Grade x Subject x Year Cells	79,466	79,330	91,221	91,221	34,495	34,495	23,445	23,445
	Panel B: Models for change in prior-year scores							
Change in mean teacher predicted VA	0.296	0.226	0.175	0.093	0.199	0.064	0.177	0.033
	(0.039)	(0.043)	(0.023)	(0.022)	(0.033)	(0.038)	(0.040)	(0.047)
	Panel C: Models for change in end-of-year scores, with controls for change in prior-year scores							
Change in mean teacher predicted VA	0.981	0.928	0.853	0.859	0.978	0.926	0.973	0.899
	(0.030)	(0.029)	(0.019)	(0.017)	(0.028)	(0.031)	(0.035)	(0.041)
Change in mean student prior year score	0.650	0.675	0.497	0.537	0.611	0.608	0.610	0.583
	(0.004)	(0.005)	(0.009)	(0.009)	(0.006)	(0.007)	(0.007)	(0.009)

Notes: See notes to CFR (2014a), Table 5. Columns 1, 3, 5, and 7 in Panel A reproduce results from that table. Even-numbered columns add school-year fixed effects. Panel B changes the dependent variable, while Panel C adds a control for the change in the prior-year score.