

Enhancing the Efficacy of Teacher Incentives through Framing: A
Field Experiment
Web Appendix

Roland G. Fryer, Jr., Steven D. Levitt, John List, and Sally Sadoff

Contents

A Appendix Figures and Tables	2
B Data Appendix	10
B.1 Analysis Sample	10
B.2 Attrition	10
B.3 Administrative Variables	10
B.4 Teacher Survey	11
C Implementation Guide	22
C.1 2010-2011	22
C.2 2011-2012	25

A Appendix Figures and Tables

Table 1: Project Timeline and Implementation Milestones

Month	Year 1	Year 2
August	Announcement of Program	—
September	ThinkLink 1 Teacher Opt-in Deadline — —	Invite Teachers to the Second Year ThinkLink 1 Teacher Opt-in Deadline Information Sessions
October	Information Sessions Loss Teachers Paid	Loss Teachers Paid —
November	ThinkLink 2	ThinkLink 2
December	—	—
January	Tax Deferred Loss Teachers Paid ThinkLink 3	Tax Deferred Loss Teachers Paid ThinkLink 3
February	—	—
March	Interim Report Provided State Testing Teacher Survey	Interim Report Provided State Testing —
April	—	Teacher Survey
May	Incentivized ThinkLink	Incentivized ThinkLink
June	Teachers Paid	Teachers Paid
July	—	—

Notes: This table presents the major implementation milestones for the two years of the experiment.

Table 2: Summary Statistics by All Treatment Arms

	Control	Ind. Loss	Team Loss	Ind. Gain	Team Gain	<i>p-val</i>
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Panel A: Year 1</i>						
Female	0.493	0.496	0.487	0.471	0.502	0.854
Black	0.396	0.454	0.391	0.296	0.423	0.299
Hispanic	0.553	0.478	0.560	0.632	0.521	0.297
Free or Reduced Lunch	0.981	0.942	0.966	0.940	0.962	0.012
Limited English Proficiency	0.137	0.084	0.096	0.110	0.153	0.804
Special Ed	0.138	0.090	0.171	0.098	0.100	0.030
Standardized Baseline Math Score	0.027	0.019	-0.087	0.212	-0.031	0.545
Standardized Baseline Reading Score	0.017	0.069	-0.291	0.116	-0.113	0.044
Teacher Value Added	10.530	6.621	14.827	14.316	13.479	0.631
Value Added Measure Missing	0.200	0.218	0.304	0.118	0.139	0.696
<i>Panel B: Year 2</i>						
Female	0.494	0.466	0.481	0.496	—	0.677
Black	0.413	0.321	0.403	0.418	—	0.443
Hispanic	0.535	0.614	0.532	0.533	—	0.482
Free or Reduced Lunch	0.948	0.941	0.946	0.960	—	0.844
Limited English Proficiency	0.167	0.169	0.219	0.091	—	0.132
Special Ed	0.117	0.095	0.136	0.100	—	0.549
Standardized Baseline Math Score	-0.201	-0.075	0.068	0.104	—	0.120
Standardized Baseline Reading Score	-0.038	0.064	0.026	0.137	—	0.703
Teacher Value Added	15.356	16.648	8.804	18.080	—	0.302
Value Added Measure Missing	0.394	0.414	0.327	0.467	—	0.879
Panel A Observations	700	597	601	525	534	
Panel A Classrooms	36	35	28	23	24	
Panel A Joint F-Test						0.002
Panel B Observations	703	780	905	553	—	
Panel B Classrooms	45	51	56	30	—	
Panel B Joint F-Test						0.024

Notes: This table presents summary statistics and balance tests for baseline observables and pretreatment Thinklink scores.

Table 3: The Effect of Treatment on ThinkLink Math Scores: Sensitivity Checks

	Limited set of covariates (1)	Exclude students missing covariates (2)	Exclude students with multiple teachers (3)	Exclude spillover classes (4)	Exclude students non-free/reduced lunch (5)
<i>Panel A: Year 1</i>					
Any Treatment	0.168 (0.072)	0.094 (0.060)	0.213 (0.081)	0.244 (0.113)	0.180 (0.072)
Pooled Loss	0.225 (0.083)	0.140 (0.066)	0.295 (0.095)	0.345 (0.119)	0.248 (0.083)
Individual Loss	0.245 (0.087)	0.144 (0.070)	0.337 (0.092)	0.386 (0.130)	0.286 (0.092)
Team Loss	0.205 (0.112)	0.141 (0.089)	0.248 (0.138)	0.313 (0.130)	0.211 (0.108)
Pooled Gain	0.098 (0.083)	0.037 (0.071)	0.128 (0.088)	0.178 (0.114)	0.093 (0.081)
Individual Gain	0.091 (0.097)	0.015 (0.079)	0.113 (0.104)	0.156 (0.125)	0.080 (0.097)
Team Gain	0.105 (0.093)	0.062 (0.088)	0.140 (0.094)	0.208* (0.119)	0.107 (0.089)
<i>p</i> -value (Loss=Gain)	0.080	0.105	0.020	0.017	0.028
Observations	2630	2408	2280	2235	2440
Students	2460	2247	2280	2126	2279
Classrooms	135	134	130	119	135
Teachers	105	104	102	104	105
	Limited set of covariates	Exclude students missing covariates	Exclude students with multiple teachers	Exclude year 1 Loss teachers	Exclude students non-free/reduced lunch
<i>Panel B: Year 2</i>					
Any Treatment	-0.006 (0.081)	-0.013 (0.074)	0.029 (0.080)	-0.003 (0.106)	0.014 (0.078)
Pooled Loss	-0.015 (0.083)	-0.012 (0.075)	0.048 (0.079)	0.013 (0.112)	0.019 (0.079)
Individual Loss	-0.039 (0.095)	-0.011 (0.086)	0.027 (0.088)	0.019 (0.119)	-0.028 (0.091)
Team Loss	0.008 (0.089)	-0.013 (0.081)	0.071 (0.087)	0.001 (0.124)	0.069 (0.085)
Pooled Gain	0.032 (0.109)	-0.019 (0.092)	-0.046 (0.107)	-0.038 (0.121)	-0.004 (0.107)
Individual Gain	0.034 (0.109)	-0.019 (0.092)	-0.047 (0.107)	-0.038 (0.121)	-0.003 (0.108)
<i>p</i> -value (Loss=Gain)	0.595	0.919	0.250	0.611	0.787
Observations	2697	2421	2391	1661	2552
Students	2543	2291	2391	1615	2406
Classrooms	153	153	147	101	153
Teachers	113	113	110	74	113

Notes: The results we report correspond to our main specification in Table 4 with estimates for year 1 in Panel A and estimates for year 2 in Panel B. Column (1) includes only indicators for treatment, school and grade, and baseline math test scores interacted with grade. Column (2) includes the same set of covariates as in column (1) and excludes students with missing baseline test scores. Column (3) includes the same covariates as in Table 4 and excludes students who were taught by multiple teachers participating in the experiment. In column (4), for year 1 we report results after excluding students in “spillover classes,” which were not incentivized but were taught by teachers receiving incentives for other classes; for year 2, we exclude students taught by teachers assigned to the Loss treatment in year 1. Column (5) includes the same set of covariates as in column (1) and excludes students who are ineligible for free or reduced-price lunch. Standard errors are reported in parentheses and are clustered on the student and teacher level.

Table 4: The Effect of Treatment on Thinking Math Scores: Leaving Out One Score at a Time

	Year 1		Year 2		Pooled	
	Loss	Gain	Loss	Gain	Loss	Gain
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Panel A: Full Sample</i>						
	0.234	0.100	0.021	0.006	0.124	0.051
	(0.080)	(0.079)	(0.079)	(0.106)	(0.056)	(0.062)
	2630	2630	2697	2697	5327	5327
<i>Panel B: Leave-Out One School</i>						
Without School 1	0.180	0.068	0.047	0.009	0.115	0.039
	(0.078)	(0.086)	(0.085)	(0.103)	(0.060)	(0.063)
	[0.111]	[0.456]	[0.503]	[0.964]	[0.689]	[0.719]
	2285	2285	2290	2290	4575	4575
Without School 2	0.257	0.109	-0.005	-0.028	0.122	0.040
	(0.075)	(0.078)	(0.080)	(0.113)	(0.055)	(0.063)
	[0.581]	[0.736]	[0.297]	[0.201]	[0.918]	[0.544]
	2424	2424	2475	2475	4899	4899
Without School 3	0.232	0.088	0.044	0.058	0.133	0.066
	(0.092)	(0.087)	(0.090)	(0.113)	(0.062)	(0.066)
	[0.900]	[0.446]	[0.430]	[0.146]	[0.628]	[0.409]
	2348	2348	2408	2408	4756	4756
Without School 4	0.214	0.096	0.018	0.048	0.114	0.065
	(0.083)	(0.084)	(0.085)	(0.119)	(0.059)	(0.070)
	[0.242]	[0.879]	[0.916]	[0.244]	[0.593]	[0.544]
	2374	2374	2443	2443	4817	4817
Without School 5	0.269	0.141	0.080	0.111	0.172	0.116
	(0.086)	(0.081)	(0.074)	(0.101)	(0.055)	(0.062)
	[0.124]	[0.166]	[0.145]	[0.064]	[0.050]	[0.028]
	2473	2473	2469	2469	4942	4942
Without School 6	0.234	0.117	-0.003	-0.016	0.112	0.052
	(0.086)	(0.082)	(0.082)	(0.106)	(0.061)	(0.064)
	[0.987]	[0.285]	[0.422]	[0.416]	[0.422]	[0.942]
	2459	2459	2525	2525	4984	4984
Without School 7	0.272	0.093	0.036	0.010	0.148	0.048
	(0.089)	(0.091)	(0.089)	(0.125)	(0.062)	(0.072)
	[0.228]	[0.862]	[0.527]	[0.919]	[0.199]	[0.916]
	2167	2167	2269	2269	4436	4436
Without School 8	0.264	0.174	-0.060	-0.136	0.096	0.035
	(0.099)	(0.090)	(0.079)	(0.108)	(0.064)	(0.068)
	[0.426]	[0.033]	[0.071]	[0.028]	[0.302]	[0.598]
	2133	2133	2204	2204	4337	4337
Without School 9	0.207	0.032	0.011	-0.019	0.103	0.004
	(0.076)	(0.079)	(0.083)	(0.109)	(0.056)	(0.064)
	[0.475]	[0.073]	[0.510]	[0.148]	[0.326]	[0.043]
	2377	2377	2493	2493	4870	4870

Notes: This table reports the effect of the treatment on ThinkLink scores, estimated separately for various subgroups in the data. Included on the right-hand side of each regression are the same set of control variables as used in Table 3. We present results both estimating years of the experiment separately and pooling across years of data. The coefficients we report are Intent-to-Treat estimates, i.e. students are classified based on their initial classroom assignment. We also report p-values from tests of equal coefficients between grade level groups, genders, races and baseline testscore groups. Standard errors are reported in parentheses and are clustered on the teacher and the student level.

Table 5: The Effect of Treatment on ThinkLink Reading Scores

	Year 1	Year 2	Pooled
	(1)	(2)	(3)
Any Treatment	0.005 (0.067)	0.091 (0.067)	0.028 (0.040)
Pooled Loss	0.058 (0.082)	0.102 (0.069)	0.053 (0.045)
Individual Loss	-0.103 (0.109)	0.095 (0.077)	0.015 (0.057)
Team Loss	0.164 (0.088)	0.109 (0.079)	0.086 (0.058)
Pooled Gain	-0.036 (0.076)	0.039 (0.088)	-0.024 (0.058)
Individual Gain	0.039 (0.092)	0.039 (0.088)	0.038 (0.066)
Team Gain	-0.175 (0.088)	—	-0.165 (0.081)
Probability(Gain=Loss)	0.269	0.394	0.221
Observations	2556	2561	5117
Clusters	142	150	292

Notes: The results we report are from regressions with standardized ThinkLink reading test scores as the outcome variable. Included on the right-hand side of the regression is the student’s treatment assignment, demographic and socio-economic characteristics of the student (gender, race, eligibility for free or reduced price lunch, Limited English Proficiency status, eligibility for Special Education services), school and grade fixed effects, and baseline reading test scores (interacted with grade). We impute missing data with zeros, adding indicator variables for missing values. We present results both estimating years of the experiment separately and pooling across years of data. We obtain our Year 2 estimates controlling for Year 1 treatment status. When pooling the data across years, the control variables are fully interacted with year dummies. We show results for each of our treatment arms separately, as well as pooling the team and individual treatments and pooling the gain and loss treatments. The coefficients we report are Intent-to-Treat estimates, i.e. students are classified based on their initial classroom assignment. Standard errors are reported in parentheses and are clustered at the classroom level. The number of observations and the number of clusters are reported at the bottom of the table.

Table 6: Teacher Attrition in Long Run Test Score Sample

	(1)	(2)	(3)	(4)	(5)	(6)
	2011/12	2012/13	2013/14	2014/15	2015/16	% Yrs Observed
Gain	0.224 (0.169)	0.131 (0.176)	0.281 (0.179)	0.252 (0.181)	0.103 (0.183)	0.198 (0.162)
Loss	0.214 (0.166)	0.079 (0.180)	0.112 (0.182)	0.104 (0.184)	-0.088 (0.177)	0.084 (0.160)
VA	0.009 (0.006)	0.002 (0.008)	0.004 (0.008)	0.004 (0.008)	0.004 (0.008)	0.005 (0.007)
Missing VA	-0.082 (0.101)	-0.032 (0.126)	-0.068 (0.127)	-0.007 (0.126)	-0.037 (0.125)	-0.045 (0.093)
VA \times Gain	-0.009 (0.006)	-0.011 (0.008)	-0.013 (0.008)	-0.014 (0.008)	-0.012 (0.008)	-0.012 (0.007)
VA \times Loss	-0.010 (0.006)	-0.007 (0.008)	-0.006 (0.009)	-0.003 (0.009)	0.000 (0.009)	-0.005 (0.007)
Constant	0.645 (0.158)	0.583 (0.158)	0.434 (0.162)	0.422 (0.164)	0.428 (0.163)	0.502 (0.151)
Observations	105	105	105	105	105	105

Notes: This table reports results from an OLS regression with being in the sample in a given year (columns 1-5) and the fraction of years in the sample (column 6) as the outcome variable. Gain and Loss indicate treatment in year 1 of the study. Value added (VA) is a teacher's average value added in the year prior to the experiment. Standard errors are clustered by teacher.

Table 7: The Long-Term Impact of Treating Teachers on their Value Added: Inverse Probability Weighted Estimates

		2010/11					
		ThinkLink		ISAT/ITBS			
		K-8		K-8			
<i>Panel A: Treatment Year</i>							
Any		0.175		0.107			
		(0.070)		(0.075)			
Loss		0.234		0.151			
		(0.080)		(0.084)			
Gain		0.100		0.048			
		(0.079)		(0.084)			
p-value (Loss=Gain)		0.05		0.17			
Observations		2630		2552			
Students		2460		2367			
Teachers		105		105			
		2011/12		2012/13	2013/14	2014/15	2015/16
		ThinkLink	ISAT	ISAT	ISAT	PARCC	PARCC
		K-8	3-8	3-8	3-8	3-8	3-8
<i>Panel B: Post-Treatment Years</i>							
Any		0.053	0.111	-0.214	0.012	0.586	0.324
		(0.096)	(0.139)	(0.194)	(0.159)	(0.201)	(0.184)
Loss		0.160	0.220	-0.092	0.083	0.831	0.175
		(0.098)	(0.131)	(0.185)	(0.174)	(0.204)	(0.205)
Gain		-0.047	0.035	-0.287	0.005	0.564	0.323
		(0.102)	(0.146)	(0.207)	(0.162)	(0.185)	(0.179)
p-value (Loss = Gain)		0.02	0.04	0.20	0.47	0.06	0.26
Observations		2115	1498	1296	1150	1079	856
Students		1973	1368	1270	1139	1078	855
Teachers		87	52	41	36	36	36
		2010/11-2015/16			2011/12-2015/16		
		ISAT/ITBS/PARCC			ISAT/PARCC		
		K-8			2-8		
<i>Panel C: Pooled</i>							
Any		0.083			0.040		
		(0.065)			(0.119)		
Loss		0.171			0.153		
		(0.066)			(0.116)		
Gain		0.006			-0.012		
		(0.074)			(0.119)		
p-value (Loss = Gain)		0.00			0.02		
Observations		8446			5894		
Students		3953			3281		
Teachers		105			65		

Notes: This table reports the effect of the treatment on ThinkLink scores standardized within year and grade, estimate separately for various subgroups in the data. Included on the right-hand side of each regression are the same set of control variables as used in Table 3. We present results both estimating years of the experiment separately and pooling across years of data. The coefficients we report are Intent-to-Treat estimates, i.e. students are classified based on their initial classroom assignment. We also report p-values from tests of equal coefficients between grade level groups, genders, races and baseline test score groups. Standard errors reported in parentheses are clustered on the teacher and student level. Observations are weighted by the inverse of the predicted probability that a teacher is in the sample in a given year. The probability is predicted using the coefficients from the regressions reported in Table A.6. student level.

Table 8: The Effect of Treatment on Math Scores: Teachers in Study Both Years

	ThinkLink			State Tests		
	Year 1	Year 2	Pooled	Year 1 ITBS/ ISAT Grades K-8	Year 2 ISAT Grades 3-8	Pooled ITBS/ ISAT Grades K-8
	(1)	(2)	(3)	(4)	(5)	(6)
Any Treatment	0.269 (0.092)	-0.043 (0.088)	0.120 (0.064)	0.147 (0.095)	-0.069 (0.110)	0.072 (0.070)
Pooled Loss	0.311 (0.109)	-0.036 (0.090)	0.134 (0.071)	0.177 (0.108)	-0.020 (0.111)	0.104 (0.079)
Individual Loss	0.414 (0.108)	-0.123 (0.109)	0.114 (0.078)	0.187 (0.112)	-0.071 (0.131)	0.081 (0.084)
Team Loss	0.228 (0.132)	0.050 (0.094)	0.150 (0.084)	0.184 (0.130)	0.062 (0.111)	0.141 (0.096)
Pooled Gain	0.227 (0.092)	-0.070 (0.130)	0.094 (0.071)	0.116 (0.100)	-0.203 (0.137)	0.024 (0.076)
Individual Gain	0.255 (0.099)	-0.051 (0.128)	0.107 (0.082)	0.030 (0.109)	-0.184 (0.137)	-0.048 (0.085)
Team Gain	0.208 (0.099)		0.075 (0.080)	0.192 (0.113)		0.135 (0.095)
<i>p</i> -value(Loss=Gain)	0.24	0.76	0.51	0.42	0.10	0.21
Observations	2122	2115	4237	2052	1498	3550
Students	2019	1973	2942	1944	1368	2488
Teachers	87	87	87	87	52	87

Notes: The results we report are from regressions with ThinkLink test scores (columns 1-3) or state test scores (ITBS for grades K-2 and ISAT for grades 3-8 in columns 4-6) as the outcome variable. Included on the right-hand side of the regression is the student's treatment assignment, demographic and socio-economic characteristics of the student (gender, race, eligibility for free or reduced price lunch, Limited English Proficiency status, eligibility for Special Education services), school and grade fixed effects, once-lagged test scores (interacted with grade), and once-lagged teacher value added. We impute missing data with zeros, adding indicator variables for missing values (missing value added measures are replaced with the sample mean). We present results both estimating years of the experiment separately and pooling across years of data. We obtain our Year 2 estimates controlling for Year 1 treatment status. When pooling the data across years, the control variables are fully interacted with year dummies. We show results for each of our treatment arms separately, as well as pooling the team and individual treatments and pooling the gain and loss treatments (we also report *p*-values from tests of equal coefficients between the loss and gain treatments). The coefficients we report are Intent-to-Treat estimates, i.e. students are classified based on their initial classroom assignment. Standard errors are reported in parentheses and are clustered on the classroom and the student level.

B Data Appendix

B.1 Analysis Sample

The sample includes all students assigned to a teacher at the beginning of the year before treatment was assigned. Students who enter the district in the middle of a school year are not included in the sample for that year. As discussed in the paper, we include all first year control students whose teacher was incentivized in other homerooms (spillover students) in the control group.

B.2 Attrition

Students in the analysis sample for whom we do not observe an end of year test score are considered attriters.

B.3 Administrative Variables

Free and Reduced Price Lunch

We construct an indicator variable set to one if a student meets guidelines for free or reduced price lunch, as determined by District administrative files, and zero otherwise.

Limited English Proficiency

We set the LEP indicator to one if the District considers a student to be a “Limited English Speaker” or a “Non English Speaker” and zero if the student is a “Fluent English Speaker” or “Native English Speaker”.

Race/Ethnicity

We code the race variables such that the four categories – white, black, Hispanic, other – are complete and mutually exclusive. Hispanic ethnicity is an absorbing state. Hence, “white” implies non-Hispanic white, “black” non-Hispanic black, and so on.

Special Education

We construct an indicator variable set to one if students receive special instruction according to an Individualized Education Plan, as recorded by the District, and zero otherwise.

Test Scores

For grades K-8, we observe scaled Math and Reading scores on four ThinkLink examinations taken during both years of the experiment. We observe baseline, fall, winter, and spring scores for grades

3-8. For a given year, teachers' rewards were determined by their students' performance on the final test (spring) of that year.

We also observe scores from state tests that students take in March of each year. Grades K-2 take the Iowa Test of Basic Skills (ITBS) and grades 3-8 take the Illinois Standard Achievement Test (ISAT). We observe math and reading scores for all grades in the first year of the experiment and only grades 3-8 in the second year of the experiment (The District did not administer the ITBS in the 2011-2012 school year). In the post-treatment analysis, we observe ISAT scores for grades 3-8 in the 2012-13 and 2013-14 school years. We observe scores for grades 3-8 on the The Partnership for Assessment of Readiness for College and Careers (PARCC) assessment in 2014-15 and for grades 2-8 on the PARCC in 2015-16.

To simplify interpretations, we normalize all scores to have mean zero and standard deviation one by year, grade, subject, and test. When test scores are used as control variables, we leave them unstandardized and use the scale score interacted with student grade.

Test Scores

Teacher Value Added is a measure in the year before the experiment begins, the 2009-2010 school year using the state tests described above. A teacher's value added is the average growth of her students on the test in percentile units.

Treatment Variables

All treatment designations are based on the teacher(s) to which a student is assigned at the beginning of a school year. When a student has multiple teachers on a given test, the student receives one observation per teacher in the analysis sample.

B.4 Teacher Survey

In March of 2011 and April of 2012, we administered a 41 question online survey to all teachers that participated in the experiment for that school year. Below we will describe the variables we used in our analysis that were generated from this survey and provide a complete list of the questions we asked in the survey.

Variables

Hours Lesson Planning

Teachers were asked “Last week, how much time did you spend on lesson planning? No time, 0-1 hour, 2-3 hours, 4-6 hours, 7-9 hours, 10-15 hours, or Over 15 hours?” and responded with one of those options. We code the responses as 0 for “No time”, as 18 for “Over 15 hours”, and as the midpoint of the two numbers shown in the other options.

Hours Grading

Teachers were asked “Last week, how much time did you spend on grading homework? No time, 0-1 hour, 2-3 hours, 4-6 hours, 7-9 hours, 10-15 hours, or Over 15 hours?” and responded with one of those options. We code the responses as 0 for “No time”, as 18 for “Over 15 hours”, and as the midpoint of the two numbers shown in the other options.

Hours Calling or Meeting w/ Parents

Teachers were asked “Last week, how much time did you spend on calling or meeting with parents? No time, 0-1 hour, 2-3 hours, 4-6 hours, 7-9 hours, 10-15 hours, or Over 15 hours?” and responded with one of those options. We code the responses as 0 for “No time”, as 18 for “Over 15 hours”, and as the midpoint of the two numbers shown in the other options.

Hours Tutoring Outside of Class

Teachers were asked “Last week, how much time did you spend on tutoring students outside of regular class time? No time, 0-1 hour, 2-3 hours, 4-6 hours, 7-9 hours, 10-15 hours, or Over 15 hours?” and responded with one of those options. We code the responses as 0 for “No time”, as 18 for “Over 15 hours”, and as the midpoint of the two numbers shown in the other options.

Hours Leading Extracurricular Activities

Teachers were asked “Last week, how much time did you spend on leading extracurricular activities after school? No time, 0-1 hour, 2-3 hours, 4-6 hours, 7-9 hours, 10-15 hours, or Over 15 hours?” and responded with one of those options. We code the responses as 0 for “No time”, as 18 for “Over 15 hours”, and as the midpoint of the two numbers shown in the other options.

Hours Completing Administrative Work

Teachers were asked “Last week, how much time did you spend on administrative work? No time, 0-1 hour, 2-3 hours, 4-6 hours, 7-9 hours, 10-15 hours, or Over 15 hours?” and responded with one of those options. We code the responses as 0 for “No time”, as 18 for “Over 15 hours”, and as the midpoint of the two numbers shown in the other options.

Hours Completing Professional Development

Teachers were asked “Last week, how much time did you spend on coursework for professional development? No time, 0-1 hour, 2-3 hours, 4-6 hours, 7-9 hours, 10-15 hours, or Over 15 hours?” and responded with one of those options. We code the responses as 0 for “No time”, as 18 for “Over 15 hours”, and as the midpoint of the two numbers shown in the other options.

Personal Money Spent on Class Materials

Teachers were asked “This year, have you spent any of your personal money on materials for your students? If so, about how much of your personal money have you spent? I have not spent my personal money, \$1-\$100, \$100-\$250, \$250-\$500, or Over \$500” and responded with one of those options. We code the responses as 0 for “I have not spent my personal money”, as 800 for “Over \$500”, and as the midpoint of the two numbers shown in the other options.

Complete List of Survey Questions

Q1 At which school do you teach?

Q2 What grade(s) do you teach?

Q3-1 Last week, how much time did you spend on lesson planning?

No time

0-1 hour

2-3 hours

4-6 hours

7-9 hours

10-15 hours

Over 15 hours

Q3-2 Last week, how much time did you spend on grading homework? Same as above.

Q3-3 Last week, how much time did you spend on calling or meeting with parents? Same as above.

Q3-4 Last week, how much time did you spend on tutoring students outside of regular class time?
Same as above.

Q3-5 Last week, how much time did you spend on leading extracurricular activities after school?
Same as above.

Q3-6 Last week, how much time did you spend on administrative work? Same as above.

Q3-7 Last week, how much time did you spend on coursework for professional development? Same as above.

Q4 Besides those listed in question 3, is there any other activity that you spent at least 1 hour on last week?

Q5 This year, have you spent any of your personal money on materials for your students? If so, about how much of your personal money have you spent?

I have not spent my personal money

\$1-\$100

\$100-\$250

\$250-\$500

Over \$500

Q6 Which of the following did you do to help prepare your students for the ISAT? Please choose all that apply:

Review content areas from the test

Have students go through sample test questions

Have students take practice tests

Give students guidance on test taking strategies

Have students attend extra tutoring (e.g., after school)

Encourage students to try hard and do well on the test

Offer students incentives or other motivations to try hard and do well on the test

My students don't take ISAT

Other [Fill in blank]

Q7 If your students attended extra tutoring to prepare for the ISAT, how many students received tutoring? Did you provide the tutoring or did they receive tutoring from someone else?

Q8 If you offered students incentives, please describe the incentives:

Q9 Which of the following did you do to help prepare your students for the Spring ThinkLink?

Review content areas from the test

Have students go through sample test questions

Have students take practice tests

Give students guidance on test taking strategies

Have students attend extra tutoring (e.g., after school)

Encourage students to try hard and do well on the test

Offer students incentives or other motivations to try hard and do well on the test

My students don't take ThinkLink

Other [Fill in blank]

Q10 If your students attended extra tutoring to prepare for the Spring ThinkLink, how many students received tutoring? Did you provide the tutoring or did they receive tutoring from someone else?

Q11 If you offered students incentives, please describe the incentives:

Q12 Is there a teacher at your schools that you work closely with?

Yes

No, I do not work with any other teachers in my school

Q13 What is the name of the teacher that you work most closely with? [Respondents only saw this question if they answered "Yes" to question 12]

Q14 Why do you work with this teacher? Please choose all that apply: [Respondents only saw this question if they answered "Yes" to question 12]

We teach the same grade

We teach the same subject

We teach the same students

The other teacher is a resource teacher

Other [Fill in blank]

Q15 Do you work with this teacher on any of the following? Please choose all that apply: [Respondents only saw this question if they answered "Yes" to question 12]

Lesson planning

Curriculum development

Co-teaching

Discuss individual students' performance

Administrative work

Other [Fill in blank]

Q16 When do you usually work with this teacher? Please choose all that apply: [Respondents only saw this question if they answered "Yes" to question 12]

Before school

During class time

During lunch

During planning periods

On early dismissal days

After school

Other [Fill in blank]

Q17 Last week, about how much time did you spend working with this teacher? [Respondents only saw this question if they answered "Yes" to question 12]

We did not work together last week

0-1 hour

1-2 hours

2-3 hours

Over 3 hours

Q18 Which of the following resources would you like to have available to you and your students? Please rank the following resources in order of importance (1 = most important, 11 = least important).

More planning time

More funding to purchase classroom materials

More technology

More enrichment activities (library, music, PE, art, etc.)

More after school/extracurricular activities

More field trips

More in school tutoring

More after school tutoring

More professional development

More opportunities to work with a master teacher

Less administrative work

Q19 Did you sign up for the teacher rewards program?

Yes

No (please provide the reason why you did not sign up in the comment box)

Q20 Are you eligible to receive rewards this year as part of the teacher rewards program?

Yes

No

I don't know

Q21 Have you spoken to other teachers about the program?

Yes

No

Q22 What is the largest reward you can receive?

\$1,000

\$2,000

\$4,000

\$8,000

I don't know

Other [Fill in blank]

Q23 Which test(s) will be used to determine your rewards? Please choose all that apply:

ISAT administered in March

ThinkLink Test administered in January

ThinkLink Test administered in March

ThinkLink Test administered in May

I don't know

Other [Fill in blank]

Q24 Which subjects will be used to determine your rewards? Please choose all that apply:

Math

Reading
Language Arts
Social Studies
Science
I don't know
Other [Fill in blank]

Q25 Please fill in the name of the homeroom which will be used to determine your reward (including your own homeroom if applicable). If you have more than 4 classes that count towards your reward, please list the first 4.

Q26 As part of the teacher rewards program, were you assigned a teammate whose performance will be used to help determine your rewards?

Yes
No
I don't know

Q27 Who is your teammate? [Respondents only saw this question if they answered "Yes" to question 26]

Q28 How does your teammate's performance affect your reward? [Respondents only saw this question if they answered "Yes" to question 26]

My reward is NOT determined by my teammate's performance

My reward is determined by the AVERAGE OF my teammate's performance and my own performance

My reward is determined by the SUM OF my teammate's performance and my own performance

I don't know

Other [Fill in blank]

Q29 Please fill in the name of the homeroom which will be used to determine YOUR TEAMMATE'S performance. If there are more than 4 homerooms that count toward your teammate's performance, please list the first 4. [Respondents only saw this question if they answered "Yes" to question 26]

Q30 Is your teammate the same as the person you work most closely with in your school in questions 13-17? [Respondents only saw this question if they answered "Yes" to question 26]

Yes

No

Not Applicable

Q31 Do you work with your teammate on any of the following? Please choose all that apply:
[Respondents only saw this question if they answered “No” or “Not Applicable” to question 30]

We don't work together

Lesson planning

Curriculum development

Co-teaching

Discuss individual student's performance

Administrative work

Other [Fill in blank]

Q32 When do you usually work with your teammate? Please choose all that apply: [Respondents only saw this question if they answered “No” or “Not Applicable” to question 30]

We don't work together

Before school

During class time

During lunch

During planning periods

On early dismissal days

After school

Other [Fill in blank]

Q33 Last week, about how much time did you spend working with your teammate? [Respondents only saw this question if they answered “No” or “Not Applicable” to question 30]

We didn't work together last week

0-1 hour

1-2 hours

2-3 hours

Over 3 hours

Q34 As part of the teacher rewards program, did you already receive a check from the University of Chicago?

Yes

No

I don't know

Q35 What was the amount on the check?

\$1,000

\$2,000

\$4,000

\$8,000

I have not received the check

Other [Fill in blank]

Q36 At the end of the year, if your actual total reward is \$3000, which of the following will happen?

I will keep my check amount, nothing else will happen

I will owe the University of Chicago \$1,000

I will owe the University of Chicago \$2,000

The University of Chicago will give me an additional reward of \$1,000

The University of Chicago will give me an additional reward of \$2,000

I don't know

Other [Fill in blank]

Q37 When did you receive the check?

In the Fall

In January

I haven't received the check yet

I don't know

Q38 How much of the money have you spent?

I have not spent any of the money

\$0-\$1,000

\$1,000-\$2,000

\$2,000-\$3,000

\$3,000-\$4,000

I don't know

Q39 Did you receive an interim report describing your students' performance from the University of Chicago?

Yes

No

Q40 How much do you expect to receive in total rewards from the Teacher Rewards program?

0-\$2,000

\$2,000-\$4,000

\$4,000-\$6,000

\$6,000-\$8,000

Q41 Who do you think are the best teachers at your school? Please list up to 3 teachers in the space provided below.

C Implementation Guide

The sections below describe the implementation for each year of the experiment in detail.

C.1 2010-2011

As specified in our agreement with the Teachers Union, program participation was made available to every K-8 classroom teacher as well as reading and math interventionists. This led to approximately 160 teachers/interventionists being eligible to participate in the experiment. After we introduced the program at the district-wide Teacher Institute Day at the start of the school year, teachers had until the end of September to opt-in to the program; 150 teachers (93.75%) elected to participate.

After receiving the list of teachers that opted-in to the experiment, participating teachers were randomly assigned to control or one of the four treatment arms. Before any randomization occurred, we paired all teachers in each school with their closest match by grade, subject(s), and students taught. From there, the randomization procedure was straightforward for all “contained” teachers – i.e. those who teach the same homeroom for the entire day. Contained teachers were randomly assigned to one of the four treatments, or the control group, subject to the restriction that teachers in the “team” treatments had to be in the same treatment group as his/her teammate. Teachers who taught multiple homerooms are subject to a slightly different procedure. Their students were grouped by class and subject and randomized into one of the treatment groups. As a result, a teacher could receive incentives based on the performance of some of her classes taught throughout the day and no incentives for others.

To improve balance among the control group and the treatment arms, over a pure random draw, we re-randomized teachers after the initial draw. First, we calculated a balance statistic for the initial assignments, defined as the sum of the inverse p -values from tests of balance across all five groups. We used chi-squared tests to test for balance across categorical variables (school, grade, and subject) and rank-sum tests for continuous variables (baseline ThinkLink math score, baseline ThinkLink reading score, percent female, percent black, percent Hispanic, and contact minutes with teacher). Our algorithm then searches for teachers to “swap” until it finds a switch that does not violate any of the rules outlined above. If switching these teachers’ treatments would improve the balance statistic, the switch is made; otherwise it is ignored. The algorithm continues until it

has tested forty potential swaps.

After teachers were randomized into groups, information sessions about the experiment and the treatment groups were held in October. At this time, teachers in Loss groups signed a contract stating that they would return the difference of \$4,000 and their final reward if their performance led to a reward less than \$4,000. For tax reasons, some Loss teachers requested that we issue the upfront payment in 2011. About half of teachers received the Loss reward at the beginning of January 2011. All other Loss teachers received a \$4,000 check in October.

Throughout the school year, four ThinkLink assessments (September, November, January, and May) were administered to all K-8 students in the Chicago Heights school district (note that grades K-2 only took the May assessment). If students were missing test scores from the previous year, the September ThinkLink scores were used to predict students' outcomes for the incentivized exam and assign students to a neighborhood of 9 similar students in order to use the "pay for percentile" methodology described below. Specifically, for each grade level, we predicted end of year scores by regressing the May ThinkLink scores from all years before 2011 onto all tests that those students took in the preceding school year and student-level demographics. For example, the regression for third graders would be a pooled regression of 2010 third grade May ThinkLink scores onto second grade test scores from the 2008-2009 school year, 2009 third grade May ThinkLink scores onto second grade test scores from the 2007-2008 school year, and so forth. Note that the regression specifications vary by grade level because all grades did not take the same set of tests. We then take the coefficients from these regressions and estimate what a student's outcome will be on the 2011 May ThinkLink given their current grade level, previous year test scores, and demographics. If students were missing previous year test scores, we attempted to do a similar procedure but instead used September ThinkLink scores instead of previous year test scores. For kindergartners, we used pretreatment standardized Developmental Reading Assessment (DRA) as their predicted outcome. If the above methods did not work (i.e. they were missing previous year and September ThinkLink test scores), we imputed their baseline score to be the mean of the standardized scores which equals zero. The January ThinkLink assessment was used to provide treatment teachers with an interim report of their students' progress in March (the results were for informational use only and did not affect teachers' final reward). Teachers in the four incentive groups received rewards based on their students' performance on the May ThinkLink assessment.

Each ThinkLink subject test lasts 30-60 minutes and is either taken on the computer (3rd-8th grade students at all but one school) or on paper (all K-2 students and 3rd-8th grade students at

one school). All students were tested in math and reading. In addition, 4th and 7th grade students took a science test as they do on the statewide tests. We proctored all end of the year testing in order to ensure consistency and discourage cheating. We used the prepackaged ThinkLink test C for all grades.

In addition to the ThinkLink assessments, students took statewide exams in March. Grades 3-8 took the math and reading subject tests of the Illinois Standard Achievement Test (ISAT). In addition, 4th and 7th graders took a science subject test. All public school students are required to take the ISAT unless they are medically excused or have a severe disability. Students with moderate disabilities or limited English proficiency must take both tests, but may be granted special accommodations (additional time, translation services, alternative assessments, and so on) if they meet the requirements set by the state. In order to ensure that as many students take the test as possible, the state provides a make-up testing window and the principal/district is required to provide the state with a written explanation of why a student registered at a specific school was not tested. Grades K-2 took the Iowa Test of Basic Skills (ITBS), developed by the University of Iowa. Since this exam was not a high-stakes exam, special accommodations and make up testing were much more relaxed than for the ISAT.

In March, we administered an online survey to all participating teachers about their time use, collaboration with fellow teachers, and knowledge about the rewards program. See Online Appendix A for a list of all the questions that were in the survey.

In June, we calculated rewards using the “pay for percentile” methodology developed by Barlevy and Neal (2011). At baseline, we placed each student in a bin with his nine nearest neighbors in terms of end of year predicted scores described above. Students are placed in separate bins for each subject. We then ranked each student within their bins according to improvement between his baseline and end of the year test scores. We rounded up students’ percentile rankings within their neighborhoods to 100%, 90%, 80%20%, and 10%. Each teacher then received an “overall percentile,” which was the average of all her students’ percentile ranks within their bins for the subject that the teacher taught. We rewarded teachers of contained classrooms based on the performance of their homeroom on both reading and math (and science in 4th and 7th grades only). We rewarded teachers of rotating classrooms on the subset of the homeroom-subjects they taught that were randomized into a treatment group. Teachers received \$80 per percentile for a maximum possible reward of \$8,000. Note that teachers in non-tested subjects (e.g. social studies and language arts) had to have the opportunity to earn incentives. However, since the district

only administered Math, Reading, and Science tests, we offered incentives to these teachers based on their students' performance on the Reading exam. Gain teachers received their final rewards in June in the form of a check. Loss teachers whose final rewards were greater than \$4,000 also received a check for the difference at this time. Loss teachers whose final rewards were less than \$4,000 were notified that they owed us money and we collected money from these teachers over the next couple of months.

See Appendix Table 2 in the main text for a timeline of the implementation described above.

C.2 2011-2012

In the second year of the experiment, program participation was limited to K-8 classroom teachers who taught math or reading. This led to 163 teachers being eligible to participate in the experiment. In early September, we sent invitations via letters and e-mail to eligible teachers to participate in the second-year of the program. Teachers had until the middle of September to opt-in to the program; 148 teachers (90.80%) elected to participate.

Participating teachers were randomly assigned to control or one of the three treatment arms (the "Gain Team" treatment group was dropped to increase power in the second year of the experiment). The randomization procedure was similar to the first year procedure except that we had to impose a constraint that no first year control teacher would end up in control again in the second year due to an agreement with the school district. Also, we included a categorical variable for new and returning teachers and only balanced on the categorical variables, not the continuous variables described above for the year one randomization. Note that we randomized at the teacher level this year for all teachers.

Information sessions were held at the end of September. As in the first year, Loss group teachers during this time. Again, some Loss teachers requested that we issue the upfront payment in 2012. About twenty percent of teachers received the Loss reward at the beginning of January 2012. All other Loss teachers received a \$4,000 check in October.

The number of and the timing of the ThinkLink assessments were identical to what we did in the first year of the experiment. In addition, we used the same methods to predict students' baseline outcomes and determine a teacher's reward based off of the May ThinkLink results. We provided both treatment and control teachers with an interim report of their students' progress in March based off January ThinkLink assessment.

In the second year of the experiment, more students took the test on the computer (3rd-8th

grade students in all schools and 2nd grade students in some schools) than the previous year. The same subjects as the first year were tested. We again proctored all end of the year testing in order to ensure consistency and discourage cheating. We used the prepackaged ThinkLink test C for grades K-2. We used ThinkLink probes that we created from a bank of questions for grades 3-8 because the district hadn't purchased test C.

Students again took statewide exams in March. Grades 3-8 took the high-stakes Illinois Standard Achievement Test (ISAT). Grades K-2 did not take a statewide exam this year because the Chicago Heights School District stopped administering the Iowa Test of Basic Skills (ITBS) after the 2010-2011 school year.

In April, we administered the same online survey as we did in year one to all participating teachers. See Online Appendix A for a list of all the questions that were in the survey.

In June, we again calculated rewards using the “pay for percentile” methodology developed by Barlevy and Neal (2011). Payments were initiated in June and balances for the Loss teachers were settled by the end of the summer.

See Appendix Table 2 in the main text for a timeline of the implementation described above.