Online Appendix

Does the "Boost for Mathematics" Boost Mathematics? A Large-Scale Evaluation of the "Lesson Study" Methodology on Student Performance

Erik Grönqvist, Björn Öckert and Olof Rosenqvist

Outline

A. Content of the Boost for Mathematics

B. Supplemental results

C. Supplemental teacher survey results

D. Reliability of test scores

E. Cost and benefit calculations

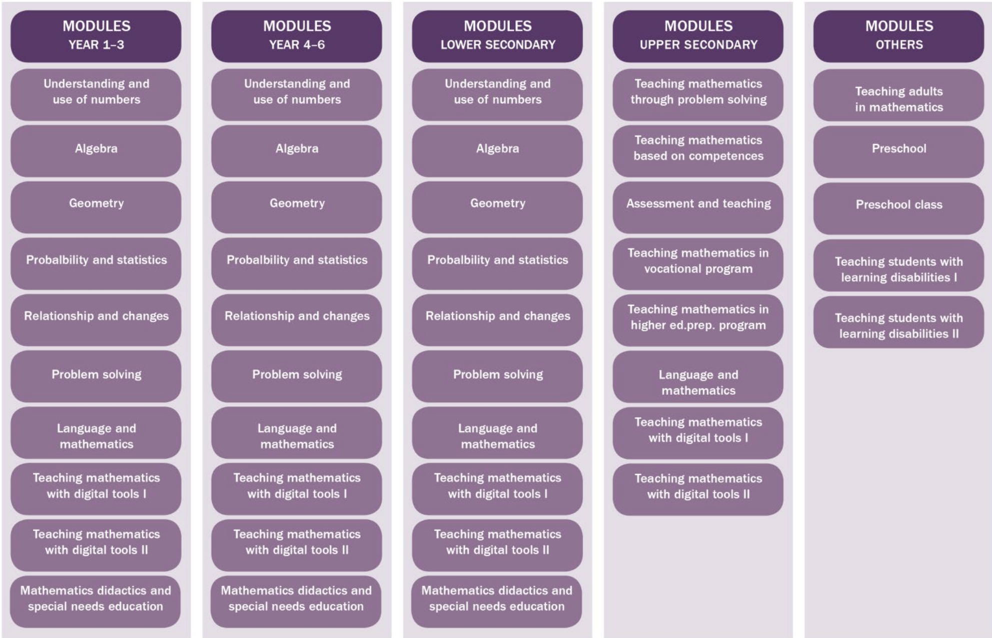
A. Content of the Boost for Mathematics

The Boost for Mathematics is based on educational modules with didactic support material (available online) covering different mathematical content. For each stage of compulsory school and upper secondary school there are separate modules, which are adapted to the didactic challenges at the specific level of schooling. Compulsory school has 10 different educational modules at each stage covering different mathematical themes; see Figure A1 for a full list of modules. All modules address the theme from the didactic perspectives: formative assessment or assessment for learning; competencies in the Swedish curriculum; classroom norms and socio-mathematical norms; interaction in the classroom (for details see Lindvall et al. 2022). There can also be additional didactic perspectives in the modules e.g., ICT, a historical perspective, or variation theory of learning.

The support material (e.g., texts, articles, films, and mathematics problems) in the modules is based on courses and syllabi, research on learning and teaching mathematics, and analyses of Swedish students' performance in national and international assessments. To ensure the quality and relevance of the didactic support material, each module is developed by two universities or teacher training colleges in collaboration, where the content is assessed by independent researchers in a peer review process. Focus groups of teachers have also been involved in this process. All modules consist of 8 parts, with each working through a learning cycle of 4 steps; see Figure A2 for a typology.

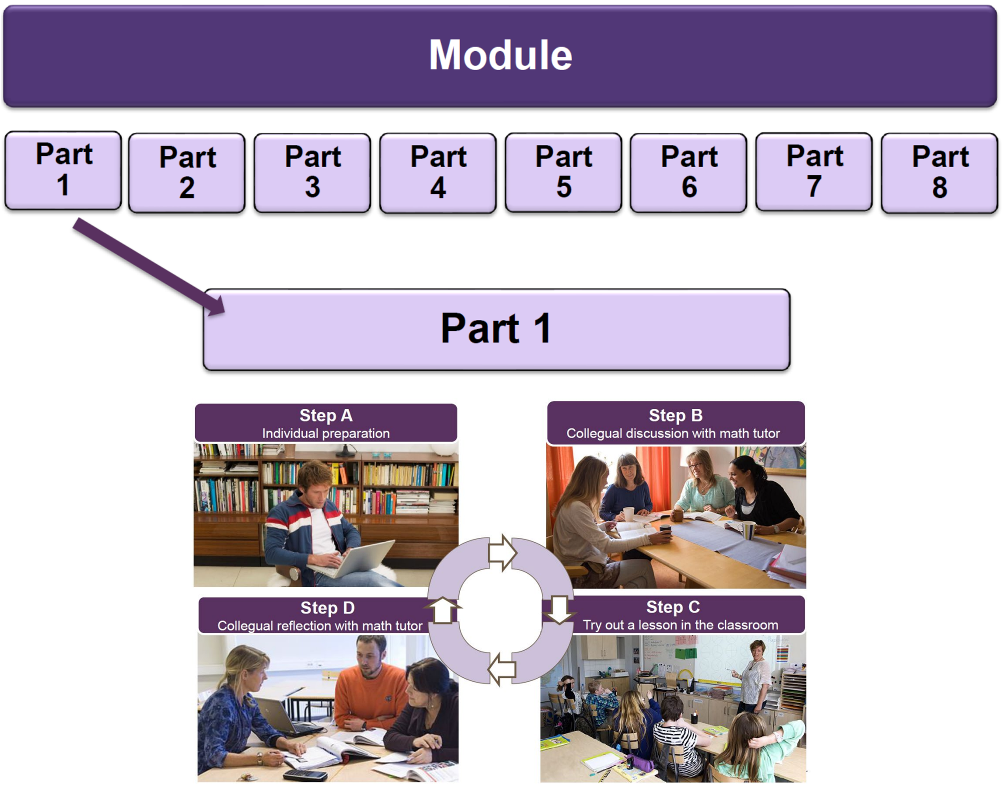
The set-up of the program is based on the local needs of the school, and it is the principal together with the tutor and teacher group – and in collaboration with the school district – that decides on which two modules to work with. The local principal is responsible for organizing the teacher groups and allocating time for training activities within the regular working hours.

Figure A1. Content of modules



Source: (Skolverket 2018)

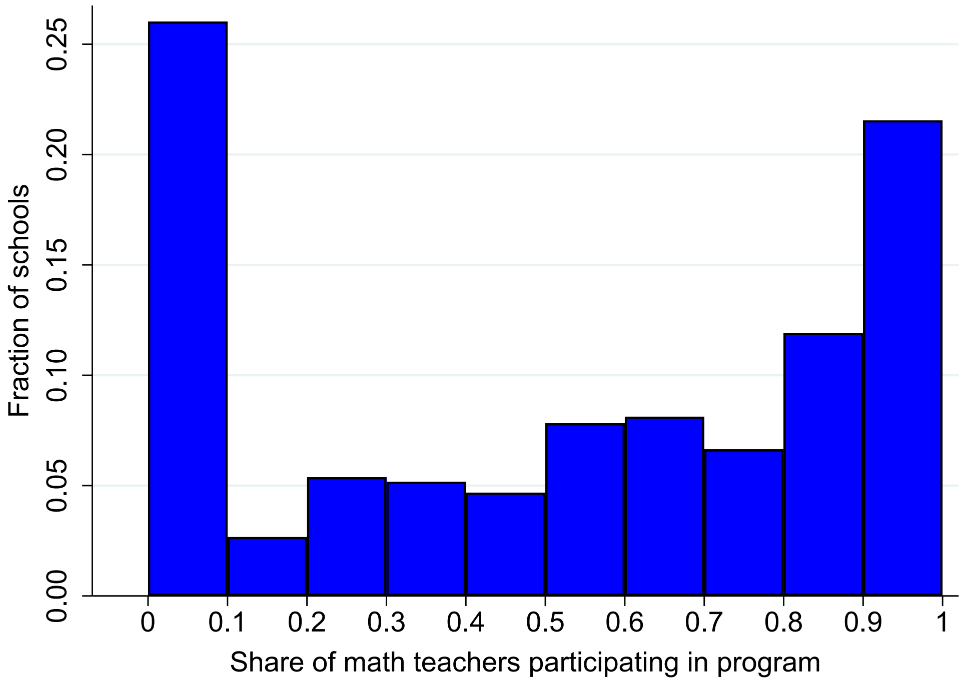
Figure A2. Illustration of the learning cycle in the module



Source: (Skolverket 2018)

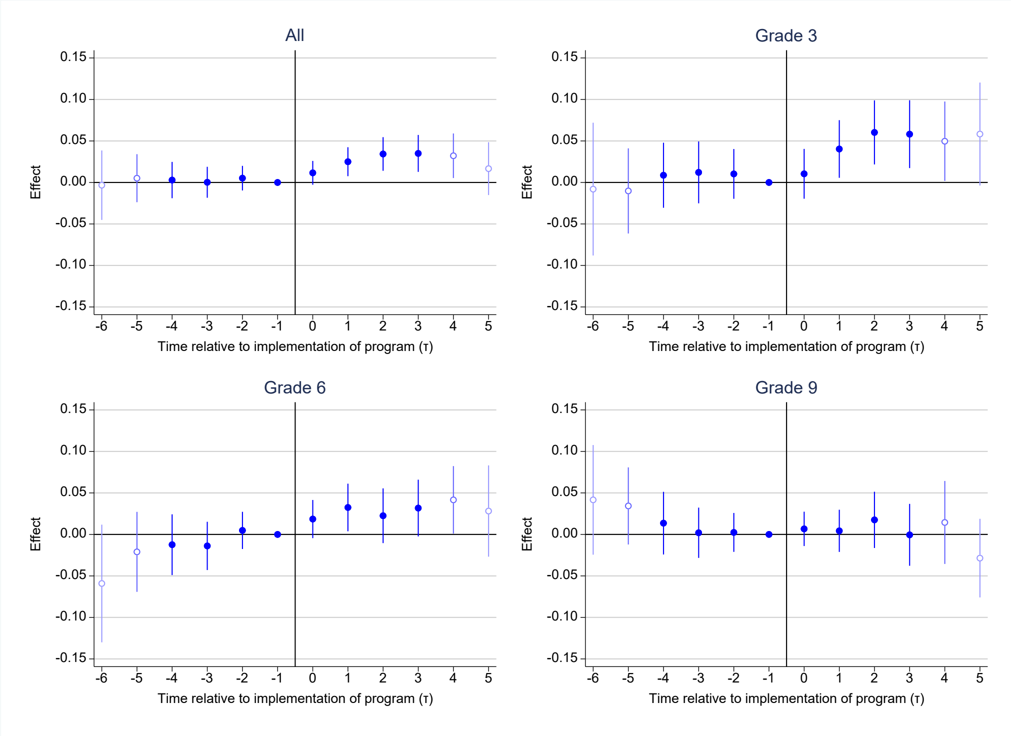
B. Supplemental results

Figure B1. Distribution of the share of participating teachers in schools



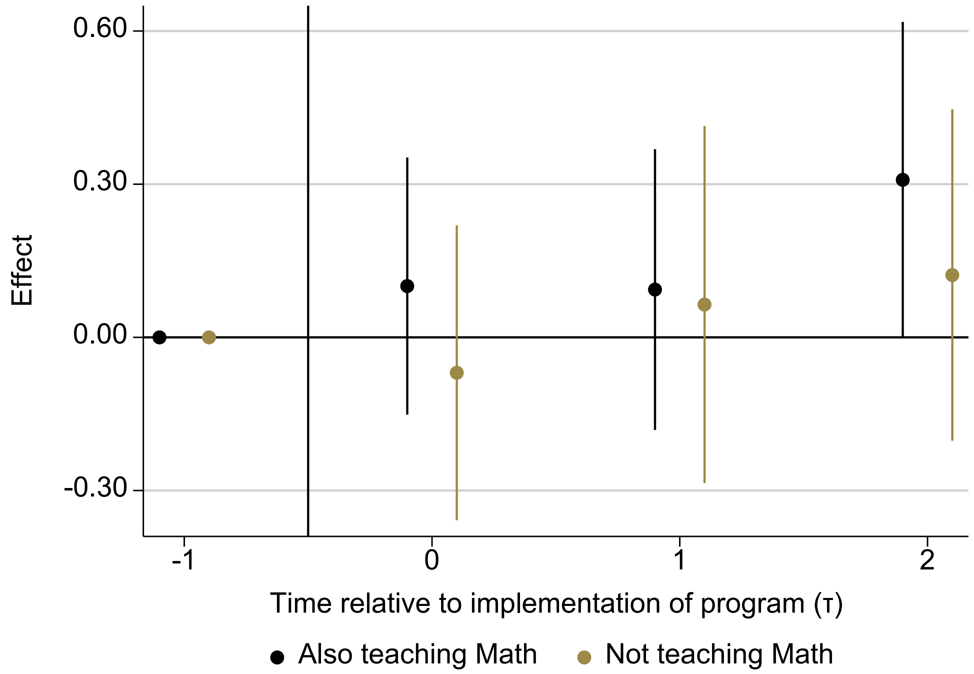
Note: The figure shows the distribution of schools with different share of mathematics teachers that receive the government grant for participating in the Boost for Mathematics.

Figure B2. Effects of the Boost for Mathematics on test scores in mathematics, by stage



Note: The figure displays reduced form effects of the Boost for Mathematics on standardized test scores in mathematics along with 95-percent confidence bands for all grades combined, and separate for grades 3, 6, 9. Estimates in a slightly lighter shade (τ = -5 and 4) are based only on schools in two waves of the intervention and estimates in the lightest shade (τ = -6 and 5) only on schools in one wave. The model includes school-by-wave fixed effects and time-by-wave fixed effects that vary by municipal and voucher schools. Test scores are measured in the end of lower/middle/higher stage (grade 3/6/9). Students are sampled in the beginning of the lower/middle/higher stage (grades 1/4/6) and assigned the treatment status of the school they are expected to attend in the end of the stage. Standard errors are clustered at the school level.

Figure B3. Effects of the Boost for Mathematics on teachers' classroom practices in Swedish (First principal component)



Note: The figure displays the effects of the Boost for Mathematics on teachers' self-reported class-room practices in Swedish along with 95-percent confidence bands. All models include school-by-wave fixed effects and time-by-wave fixed effects that vary by municipal and voucher schools. The outcome variable is the first principal component of responses to the survey question: "How often do you do the following/ask students to do the following in your teaching about and with texts…?" for 31 different activities. Answers are reported as very often, often, sometimes, rarely, never, and we standardize the answers on each question. The activities are listed in the online Appendix F. Standard errors are clustered at the school level.

Table B1. Effects of the Boost for Mathematics on mathematics test scores. Alternative treatment definitions

Column:	(1)	(2)	(3)
Treatment cutoff:	0.50	0.20	0.80
All years pooled	0.0262***	0.0248***	0.0245**
	(0.0085)	(0.0084)	(0.0096)
School×Wave FE	Yes	Yes	Yes
Year×Private×Wave FE	Yes	Yes	Yes
Observations	2,872,984	3,259,049	2,252,445

Note: The table shows reduced form effects of the Boost for Mathematics on standardized test scores in mathematics for different definitions of schools' treatment status. The treatment cutoff values indicated in the column heading is the lowest share of mathematics teachers participating in the program required for the school to be defined as treated. All models include school-by-wave fixed effects and time-by-wave fixed effects that vary by municipal and voucher schools. Test scores are measured in the end of lower/middle/higher stage (grade 3/6/9). Students are sampled in the beginning of the lower/middle/higher stage (grades 1/4/6) and assigned the treatment status of the school they are expected to attend in the end of the stage. Cluster-adjusted standard errors at the school level are in parentheses and */**/*** refers to statistical significance at the 10/5/1 percent level.

Table B2. Effects of the Boost for Mathematics on actual exposure to the program (first stage)

Column:	(1)	(2)	(3)	(4)
Grades:	3, 6 and 9	3	6	9
All years pooled	0.8444***	0.9012***	0.8395***	0.7697***
	(0.0039)	(0.0043)	(0.0061)	(0.0077)
School×Wave FE	Yes	Yes	Yes	Yes
Year×Private×Wave FE	Yes	Yes	Yes	Yes
Observations	2,872,984	1,053,814	967,568	851,602

Note: The table shows reduced form effects of the student's expected exposure to the Boost for Mathematics on actual exposure. The outcome variable is years of exposure to the program in the school the student attends in the end of lower/middle/higher stage (grade 3/6/9). Students are sampled in the beginning of the lower/middle/higher stage (grades 1/4/6) and assigned the years of exposure to the program in the school they are expected to attend in the end of the stage. All models include school-by-wave fixed effects and time-by-wave fixed effects that vary by municipal and voucher schools. The sample studied is indicated in the column heading. Cluster-adjusted standard errors at the school level are in parentheses and */**/*** refers to statistical significance at the 10/5/1 percent level.

Table B3. Effects of the Boost for Mathematics on test scores in mathematics. Alternative levels of cluster-adjusted standard errors.

Column:	(1)	(2)	(3)
	Panel A. Separately for different years		
Implementation year	0.0116	0.0116	0.0116
	0.0073	(0.0072)	(0.0075)
1 year after implementation	0.0250***	0.0250***	0.0250***
	(0.0089)	(0.0084)	(0.0090)
2 years after implementation	0.0343***	0.0343***	0.0343***
	(0.0103)	(0.0115)	(0.0105)
3 years after implementation	0.0350***	0.0350***	0.0350***
	(0.0113)	(0.0130)	(0.0113)
4 years after implementation	0.0322**	0.0322**	0.0322**
	(0.0137)	(0.0149)	(0.0134)
5 years after implementation	0.0167	0.0167	0.0167
	(0.0162)	(0.0172)	(0.0165)
	Panel B. All years pooled		
All years	0.0262***	0.0262***	0.0262***
	(0.0085)	(0.0093)	(0.0086)
School×Wave FE	Yes	Yes	Yes
Year×Private×Wave FE	Yes	Yes	Yes
Cluster level	School	School district	School × stage
Observations	2,872,981	2,872,981	2,872,981

Note: The table shows reduced form effects of the Boost for Mathematics on test scores in mathematics using cluster-adjusted standard errors at different levels. All models include school-by-wave fixed effects and time-by-wave fixed effects that are allowed to vary by municipal and voucher schools. Test scores are measured in the end of lower/middle/higher stage (grade 3/6/9). Students are sampled in the beginning of the lower/middle/higher stage (grades 1/4/6) and assigned the treatment status of the school they are expected to attend in the end of the stage. */**/*** refers to statistical significance at the 10/5/1 percent level.

Table B4. Effects of the Boost for Mathematics on test scores in mathematics by implementation wave

Column:	(1)	(2)	(3)	(4)
Implementation wave:	2013	2014	2015	p-value test
	Panel A. Separately for different years			
Implementation year	0.0063	0.0066	0.0252	0.7076
	(0.0155)	(0.0169)	(0.0180)	
1 year after	0.0156	0.0248	0.0381**	0.7395
	(0.0169)	(0.0166)	(0.0189)	
2 years after	0.0276*	0.0250*	0.0661***	0.2681
	(0.0151)	(0.0144)	(0.0232)	
3 years after	0.0222	0.0494**	0.0395**	0.5455
	(0.0167)	(0.0211)	(0.0190)	
4 years after	0.0468**	0.0221	N/A	0.3691
	(0.0215)	(0.0175)		
5 years after	0.0167	N/A	N/A	
	(0.0162)			
	Panel B. All years pooled			
0–3 years after implementation	0.0181	0.0248*	0.0403**	0.5713
	(0.0131)	(0.0130)	(0.0162)	
0–4 years after implementation	0.0222*	0.0242*	N/A	0.9123
	(0.0134)	(0.0130)		
0–5 years after implementation	0.0213	N/A	N/A	
	(0.0132)			
School×Wave FE	Yes	Yes	Yes	
Year×Private×Wave FE	Yes	Yes	Yes	
Observations	1,030,064	1,005,036	837,881	

Note: The table shows reduced form effects of the Boost for Mathematics on test scores in mathematics separately by implementation wave using cluster-adjusted standard errors at different levels. All models include school-by-wave fixed effects and time-by-wave fixed effects that are allowed to vary by municipal and voucher schools. Column headings 1-3 indicate wave and column 4 the p-value of testing the null hypothesis of equal effects across waves. Test scores are measured in the end of lower/middle/higher stage (grade 3/6/9). Students are sampled in the beginning of the lower/middle/higher stage (grades 1/4/6) and assigned the treatment status of the school they are expected to attend in the end of the stage. */**/*** refers to statistical significance at the 10/5/1 percent level.

Table B5. Effects of the Boost for Mathematics on test scores in mathematics, by stage

Column:	(1)	(2)	(3)	(4)
Grades:	3, 6 and 9	3	6	9
Implementation year	0.0116	0.0104	0.0185	0.0067
	(0.0073)	(0.0153)	(0.0117)	(0.0105)
1 year after implementation	0.0250***	0.0403**	0.0325**	0.0043
	(0.0089)	(0.0177)	(0.0146)	(0.0130)
2 years after implementation	0.0343***	0.0603***	0.0226	0.0175
	(0.0103)	(0.0197)	(0.0168)	(0.0173)
3 years after implementation	0.0350***	0.0582***	0.0318*	-0.0005
	(0.0113)	(0.0208)	(0.0174)	(0.0190)
4 years after implementation	0.0322**	0.0497**	0.0416**	0.0144
	(0.0137)	(0.0244)	(0.0207)	(0.0255)
5 years after implementation	0.0167	0.0583*	0.0281	-0.0286
	(0.0162)	(0.0316)	(0.0280)	(0.0241)
School×Wave FE	Yes	Yes	Yes	Yes
Year×Private×Wave FE	Yes	Yes	Yes	Yes
Observations	2,872,984	1,053,814	967,568	851,602

Note: The table shows reduced form effects of the Boost for Mathematics on standardized test scores in mathematics, by stage. All models include school-by-wave fixed effects and time-by-wave fixed effects that vary by municipal and voucher schools. The sample studied is indicated in the column heading. Test scores are measured in the end of lower/middle/higher stage (grade 3/6/9). Students are sampled in the beginning of the lower/middle/higher stage (grades 1/4/6) and assigned the treatment status of the school they are expected to attend in the end of the stage. Cluster-adjusted standard errors at the school level are in parentheses and */**/*** refers to statistical significance at the 10/5/1 percent level.

Table B6 Specification tests

Column:	(1)	(2)	(3)	(4)
Outcome:	Predicted test scores	Test scores	Test scores	Test scores
Implementation year	0.0014	0.0101	0.0110	0.0094
	(0.0014)	(0.0071)	(0.0073)	(0.0071)
1 year after implementation	-0.0012	0.0262***	0.0240***	0.0253***
	(0.0018)	(0.0086)	(0.0089)	(0.0087)
2 years after implementation	0.0002	0.0341***	0.0319***	0.0319***
	(0.0020)	(0.0100)	(0.0104)	(0.0101)
3 years after implementation	0.0001	0.0350***	0.0325***	0.0325***
	(0.0024)	(0.0109)	(0.0115)	(0.0111)
4 years after implementation	0.0003	0.0319**	0.0300**	0.0300**
	(0.0030)	(0.0134)	(0.0139)	(0.0135)
5 years after implementation	0.0003	0.0163	0.0155	0.0151
	(0.0040)	(0.0159)	(0.0165)	(0.0161)
School×Wave FE	Yes	Yes	Yes	Yes
Year×Private×Wave FE	Yes	Yes	Yes	Yes
Student controls	No	Yes	No	Yes
School intervention controls	No	No	Yes	Yes
Observations	2,872,984	2,872,984	2,872,984	2,872,984

Note: The table shows reduced form effects of the Boost for Mathematics on predicted test scores and test scores in mathematics. All models include school-by-wave fixed effects and time-by-wave fixed effects that vary by municipal and voucher schools. The outcome variable is indicated in the column heading. Predicted test scores are used as student control. The school intervention controls are dummy variables for the schools' participation in the Boost for Reading, Career teachers, Teachers' salary boost and the reintroduction of the Boost for Mathematics in 2017. Outcomes are measured in the end of lower/middle/higher stage (grade 3/6/9). Students are sampled in the beginning of the lower/middle/higher stage (grades 1/4/6) and assigned the treatment status of the school they are expected to attend in the end of the stage. Cluster-adjusted standard errors at the school level are in parentheses and */**/*** refers to statistical significance at the 10/5/1 percent level.

Table B7. Effects of the Boost for Mathematics on test-taking, by quartile of predicted test scores

Column:	(1)	(2)	(3)	(4)
Sample:	P0–P25	P25–P50	P50–P75	P75–P100
A. Level				
Between and within schools	-0.0004	0.0014	0.0007	-0.0001
	(0.0056)	(0.0054)	(0.0052)	(0.0052)
B. Relative				
Between and within schools	REF	0.0017	0.0011	0.0003
		(0.0032)	(0.0038)	(0.0043)
Within schools	REF	0.0022	0.0018	0.0029
		(0.0029)	(0.0031)	(0.0032)
Observations	781,886	764,551	761,425	770,835
Mean of outcome	0.8926	0.9411	0.9511	0.9487

Note: The table shows reduced form effects of the Boost for Mathematics on the probability to take the standardized test in mathematics, divided by quartile of students' predicted test scores. In all models, treatment is interacted by quartile of students predicted test scores. Between-and-within school models include school-by-wave fixed effects and wave-by-time-by-quartile fixed effects that vary by municipal and voucher schools. Within-school models include wave-by-school-by-time fixed effects and wave-by-time-by-quartile fixed effects that vary by municipal and voucher schools. Panel A show results in levels and Panel B relative to P0-P25. Column headings indicate quartile in the distribution of predicted test scores. The outcome is measured in the end of lower/middle/higher stage (grade 3/6/9). Students are sampled in the beginning of the lower/middle/higher stage (grades 1/4/6) and assigned the treatment status of the school they are expected to attend in the end of the stage. Cluster-adjusted standard errors at the school level are in parentheses and */**/*** refers to statistical significance at the 10/5/1 percent level.

Table B8. Effects of the Boost for Mathematics on test scores in mathematics, by immigration status and gender

Column:	(1)	(2)	(3)	(4)
Sample:	Native	Immigrant	Girls	Boys
All years pooled	0.0276***	0.0097	0.0324***	0.0207**
	(0.0085)	(0.0212)	(0.0093)	(0.0101)
School×Wave FE	Yes	Yes	Yes	Yes
Year×Private×Wave FE	Yes	Yes	Yes	Yes
Observations	2,656,286	216,698	1,405,191	1,467,793

Note: The table shows reduced form effects of the Boost for Mathematics on standardized test scores in mathematics, divided by immigration status and gender. All models include school-by-wave fixed effects and time-by-wave fixed effects that are allowed to vary by municipal and voucher schools. The sample studied is indicated in the column heading. Test scores are measured in the end of lower/middle/higher stage (grade 3/6/9). Students are sampled in the beginning of the lower/middle/higher stage (grades 1/4/6) and assigned the treatment status of the school they are expected to attend in the end of the stage. Cluster-adjusted standard errors at the school level are in parentheses and */**/*** refers to statistical significance at the 10/5/1 percent level.

Table B9. Effects of the Boost for Mathematics on test scores in mathematics for natives, by quartile of predicted test scores

Column:	(1)	(2)	(3)	(4)
Sample:	P0–P25	P25–P50	P50–P75	P75–P100
A. Relative				
Between and within schools	0.0011	0.0331***	0.0329***	0.0291***
	(0.0131)	(0.0098)	(0.0093)	(0.0104)
B. Level				
Between and within schools	REF	0.0320***	0.0318***	0.0280*
		(0.0112)	(0.0114)	(0.0143)
Within schools	REF	0.0313***	0.0306***	0.0285**
		(0.0099)	(0.0102)	(0.0134)
Observations	574,172	681,836	695,166	705,112

Note: The table shows reduced form effects of the Boost for Mathematics on standardized test scores in mathematics for natives, divided by quartile of students' predicted test scores. The quartiles are defined in the full population (natives and immigrants). In all models, treatment is interacted by quartile of students predicted test scores. Between-and-within school models include school-by-wave fixed effects and wave-by-time-by-quartile fixed effects that vary by municipal and voucher schools. Within-school models include wave-by-school-by-time fixed effects and wave-by-time-by-quartile fixed effects that vary by municipal and voucher schools. Panel A show results in levels and Panel B relative to P0-P25. Column headings indicate quartile in the distribution of predicted test scores. Test scores are measured in the end of lower/middle/higher stage (grade 3/6/9). Students are sampled in the beginning of the lower/middle/higher stage (grades 1/4/6) and assigned the treatment status of the school they are expected to attend in the end of the stage. Cluster-adjusted standard errors at the school level are in parentheses and */**/*** refers to statistical significance at the 10/5/1 percent level.

Table B10. Effects of the Boost for Mathematics on test scores in mathematics, by teacher qualifications

Column:	(1)	(2)	(3)	(4)
Characteristic:	Experience		Mathematics degree	
Sample:	High average	Low average	High share	Low share
All years pooled	0.0317***	0.0239*	0.0274**	0.0284**
	(0.0119)	(0.0137)	(0.0119)	(0.0135)
School×Wave FE	Yes	Yes	Yes	Yes
Year×Private×Wave FE	Yes	Yes	Yes	Yes
Observations	1,286,970	1,376,962	1,284,381	1,379,551

Note: The table shows reduced form effects of the Boost for Mathematics on standardized test scores in mathematics, divided by the schools' average teacher characteristics. All models include school-by-wave fixed effects and time-by-wave fixed effects that are allowed to vary by municipal and voucher schools. The sample studied is indicated in the column heading. Test scores are measured in the end of lower/middle/higher stage (grade 3/6/9). Students are sampled in the beginning of the lower/middle/higher stage (grades 1/4/6) and assigned the treatment status of the school they are expected to attend in the end of the stage. Cluster-adjusted standard errors at the school level are in parentheses and */**/*** refers to statistical significance at the 10/5/1 percent level.

Table B11. Effects of the Boost for Mathematics on test scores in mathematics, by school characteristics

Column:	(1)	(2)	(3)	(4)
Characteristic:	School size		School market	
Sample:	Small	Big	Big city	Smaller city
All years pooled	0.0164	0.0414***	0.0509***	0.0123
	(0.0112)	(0.0143)	(0.0141)	(0.0104)
School×Wave FE	Yes	Yes	Yes	Yes
Year×Private×Wave FE	Yes	Yes	Yes	Yes
Observations	1,477,846	1,270,172	1,121,355	1,751,629

Note: The table shows reduced form effects of the Boost for Mathematics on standardized test scores in mathematics, divided by school characteristics. All models include school-by-wave fixed effects and time-by-wave fixed effects that are allowed to vary by municipal and voucher schools. The sample restriction is indicated in the column heading. Test scores are measured in the end of lower/middle/higher stage (grade 3/6/9). Students are sampled in the beginning of the lower/middle/higher stage (grades 1/4/6) and assigned the treatment status of the school they are expected to attend in the end of the stage. Cluster-adjusted standard errors at the school level are in parentheses and */**/*** refers to statistical significance at the 10/5/1 percent level.

Table B12. Effects of the Boost for Mathematics on test scores in Swedish, by stage

Column:	(1)	(2)	(3)	(4)
Grades:	3, 6 and 9	3	6	9
	Panel A. Separately for different years			
Implementation year	0.0084	0.0029	0.0030	0.0179
	(0.0063)	(0.0103)	(0.0109)	(0.0112)
1 year after implementation	0.0148*	0.0106	0.0156	0.0205
	(0.0081)	(0.0130)	(0.0139)	(0.0139)
2 years after implementation	0.0178*	0.0312**	0.0059	0.0253
	(0.0093)	(0.0143)	(0.0163)	(0.0179)
3 years after implementation	0.0248**	0.0301**	0.0256	0.0076
	(0.0099)	(0.0144)	(0.0182)	(0.0178)
4 years after implementation	0.0265**	0.0393**	0.0193	0.0389
	(0.0125)	(0.0176)	(0.0222)	(0.0271)
5 years after implementation	0.0324**	0.0445*	0.0379	0.0246
	(0.0153)	(0.0243)	(0.0282)	(0.0259)
		Panel B. All years pooled		
All years	0.0188**	0.0236**	0.0153	0.0207
	(0.0077)	(0.0118)	(0.0136)	(0.0130)
School×Wave FE	Yes	Yes	Yes	Yes
Year×Private×Wave FE	Yes	Yes	Yes	Yes
Observations	2,921,877	1,054,511	983,590	883,776

Note: The table shows reduced form effects of the Boost for Mathematics on standardized test scores in Swedish, divided by stage. All models include school-by-wave fixed effects and time-by-wave fixed effects that are allowed to vary by municipal and voucher schools. The sample studied is indicated in the column heading. Test scores are measured in the end of lower/middle/higher stage (grade 3/6/9). Students are sampled in the beginning of the lower/middle/higher stage (grades 1/4/6) and assigned the treatment status of the school they are expected to attend in the end of the stage. Cluster-adjusted standard errors at the school level are in parentheses and */**/*** refers to statistical significance at the 10/5/1 percent level.

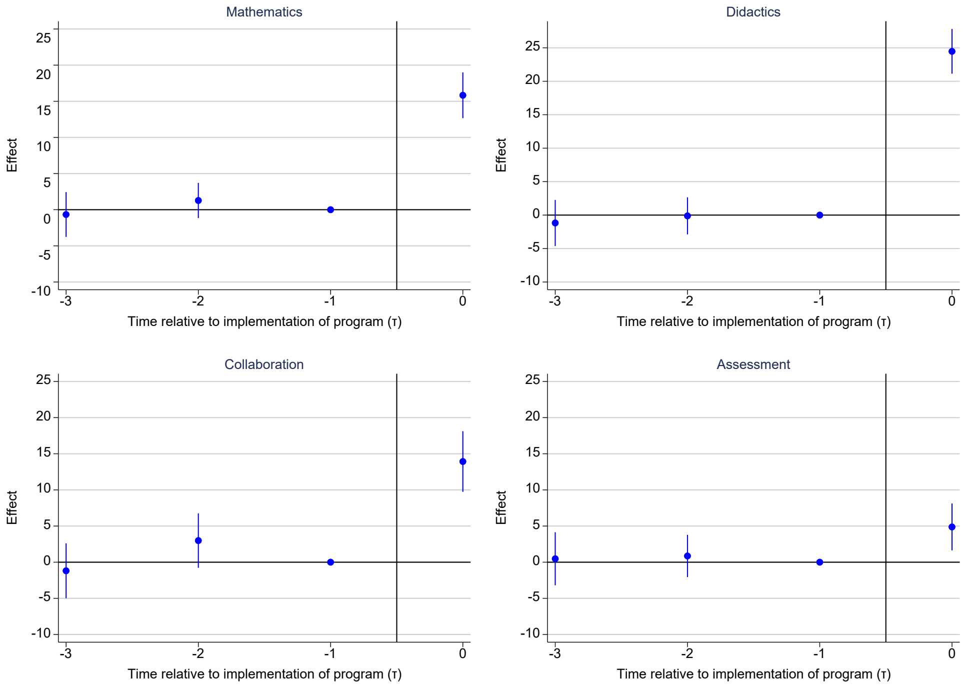
Table B 13. Effects of the Boost for Mathematics on sub-test scores in Swedish

Column:	(1)	(2)	(3)
Sub-test:	Verbal	Read	Write
Grade:	All	All	All
All years pooled	0.0117	0.0161**	0.0167**
	(0.0080)	(0.0070)	(0.0084)
School×Wave FE	Yes	Yes	Yes
Year×Private×Wave FE	Yes	Yes	Yes
Observations	2,878,650	2,900,153	2,870,713

Note: The table shows reduced form effects of the Boost for Mathematics on standardized tests in verbal proficiency, reading ability and writing skills for all stages. All models include school-by-wave fixed effects and time-by-wave fixed effects that are allowed to vary by municipal and voucher schools. The outcome studied is indicated in the column heading. Test scores are measured in the end of lower/middle/higher stage (grade 3/6/9). Students are sampled in the beginning of the lower/middle/higher stage (grades 1/4/6) and assigned the treatment status of the school they are expected to attend in the end of the stage. Cluster-adjusted standard errors at the school level are in parentheses and */**/*** refers to statistical significance at the 10/5/1 percent level.

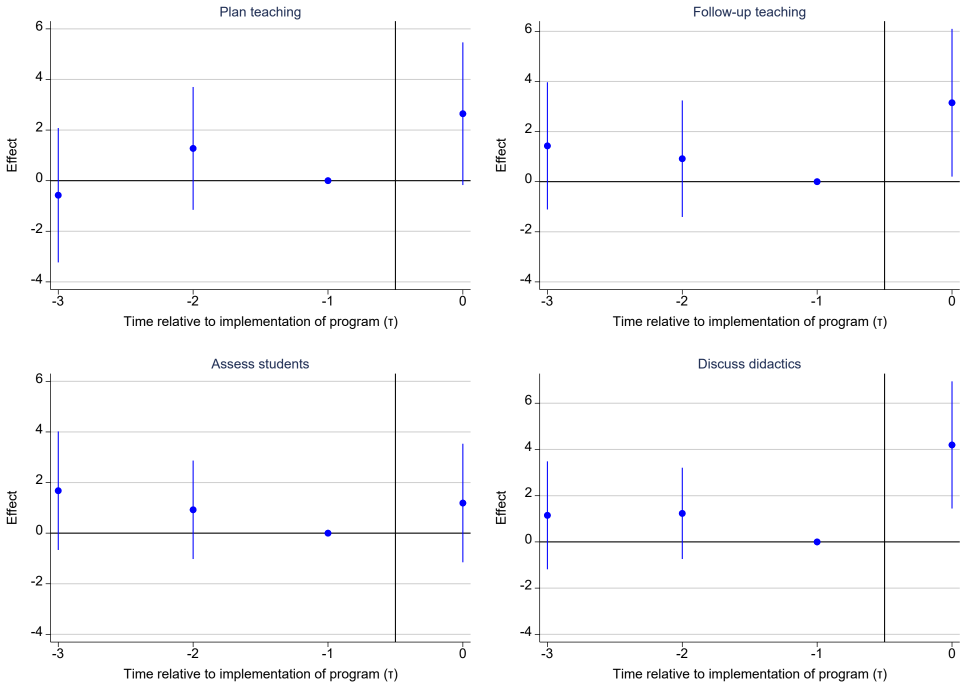
C. Supplemental teacher survey results

Figure C1. Effects of the Boost for Mathematics on teachers' training activities (hours per school year) for the third implantation wave (g = 2015)



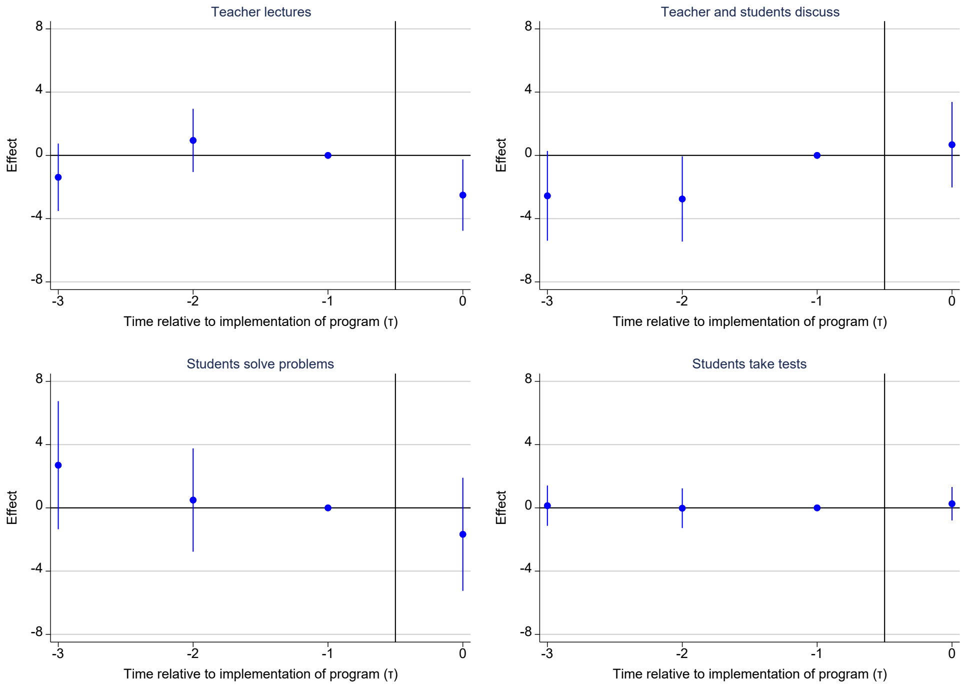
Note: The figure shows effects of the Boost for Mathematics on teachers' self-reported training activities along with 95-percent confidence bands. All models include school-by-wave fixed effects and time-by-wave fixed effects that vary by municipal and voucher schools. The outcome variable indicated in the subtitles is the answer to the survey question: "This academic year, how many hours have you participated in in-service training or other activities that involved; (a) subject knowledge in mathematics, (b) didactics of mathematics, (c) peer collaboration, or (d) student assessment?". Answers are reported as hours per school year. The table reports the p-value of the F-test for the hypothesis that all model parameters are zero.

Figure C2. Effects of the Boost for Mathematics on teacher peer collaboration activities (frequency per term) for the third implantation wave (g = 2015)



Note: The table shows effects of the Boost for Mathematics on teachers' self-reported peer collaboration activities along with 95-percent confidence bands. All models include school-by-wave fixed effects and time-by-wave fixed effects that vary by municipal and voucher schools. The outcome variable indicated in the subtitles is the answer to the survey question: "How often do you, together with another mathematics teacher; (a) plan teaching, (b) follow up on teaching, (c) follow up students' knowledge, (d) discuss instructional practices?". Answers are reported as frequency per term. The table reports the p-value of the F-test for the hypothesis that all model parameters are zero.

Figure C3. Effects of the Boost for Mathematics on teachers' classroom practices (share of lecture time) for the third implantation wave (g = 2015)



Note: The table shows effects of the Boost for Mathematics on teachers' self-reported classroom practices along with 95-percent confidence bands. All models include school-by-wave fixed effects and time-by-wave fixed effects that vary by municipal and voucher schools. The outcome variable indicated in the subtitles is the answer to the survey question: "In a typical week, what percentage of the lesson time in mathematics do students spend on each of the following activities; (a) listening to lecture-style presentations, (b) discussing problem-solving strategies together with the teacher, (c) working problems on their own or in group, (d) taking tests or quizzes?" Answers are reported as percent of time. The table reports the p-value of the F-test for the hypothesis that all model parameters are zero.

Table C1. Effects of the Boost for Mathematics on pre-determined characteristics among teachers who responded to the survey

Column:	(1)	(2)	(3)	(4)
Outcome:	Years of experience	Hours of teaching per week	Teacher diploma	University semesters in mathematics
	Panel A. Separately for different years			
Implementation year	0.71	-0.02	0.01	0.01
	(0.47)	(0.17)	(0.01)	(0.05)
1 year after implementation	-0.70	0.18	0.00	-0.02
	(0.75)	(0.24)	(0.02)	(0.09)
2 years after implementation	-0.27	0.11	-0.01	-0.06
	(1.07)	(0.36)	(0.03)	(0.15)
	Panel B. All years pooled			
All years	0.04	0.07	0.00	-0.01
	(0.60)	(0.20)	(0.01)	(0.08)
School×Wave FE	Yes	Yes	Yes	Yes
Year×Private×Wave FE	Yes	Yes	Yes	Yes
Observations	8,360	8,163	8,382	8,311
Mean of dependent var	15.27	5.58	0.95	1.76

Note: The table shows effects of the Boost for Mathematics on pre-determined characteristics among teachers who responded to the survey. All models include school-by-wave fixed effects and time-by-wave fixed effects that vary by municipal and voucher schools. Cluster-adjusted standard errors at the school level are in parentheses and */**/*** refers to statistical significance at the 10/5/1 percent level.

Table C2. Effects of the Boost for Mathematics on teachers' training activities (hours per school year)

Column:	(1)	(2)	(3)	(4)	(5)
Outcome:	Mathe-matics	Didactics	Coaching	Collabo-ration	Assess-ment
Implementation year	14.47***	23.06***	10.89***	12.36***	2.91***
	(0.79)	(0.88)	(0.70)	(0.90)	(0.85)
1 year after implementation	2.96***	4.70***	1.35**	1.78	-2.41*
	(1.05)	(1.31)	(0.58)	(1.38)	(1.23)
2 years after implementation	1.58	1.08	0.04	1.86	-2.42
	(1.53)	(1.87)	(0.82)	(1.94)	(1.77)
School×Wave FE	Yes	Yes	Yes	Yes	Yes
Year×Private×Wave FE	Yes	Yes	Yes	Yes	Yes
F-test for $\theta_0=\theta_1=\theta_2=0$	0.0000	0.0000	0.0000	0.0000	0.0000
Observations	8,373	8,373	8,373	8,373	8,373
Pre-reform mean	4.32	5.53	1.88	13.20	10.81

Note: The table shows effects of the Boost for Mathematics on teachers' self-reported training activities. All models include school-by-wave fixed effects and time-by-wave fixed effects that vary by municipal and voucher schools. The outcome variable indicated in the column heading is the answer to the survey question: "This academic year, how many hours have you participated in in-service training or other activities that involved; (1) subject knowledge in mathematics, (2) didactics of mathematics, (3) support by a coach, (4) peer collaboration, or (5) student assessment?". Answers are reported as hours per school year. The table reports the p-value of the F-test for the hypothesis that all model parameters are zero. Cluster-adjusted standard errors at the school level are in parentheses and */**/*** refers to statistical significance at the 10/5/1 percent level.

Table C3. Effects of the Boost for Mathematics on teacher peer collaboration activities (frequency per term)

Column:	(1)	(2)	(3)	(4)	(5)
Outcome:	Plan teaching	Follow-up teaching	Assess students	Discuss didactics	Classroom visits
Implementation year	3.35***	2.68***	0.80*	4.38***	0.11
	(0.56)	(0.53)	(0.47)	(0.52)	(0.35)
1 year after implementation	0.98	0.60	0.45	1.13	0.08
	(0.88)	(0.80)	(0.73)	(0.76)	(0.60)
2 years after implementation	1.67	1.55	1.59	1.76	-0.38
	(1.20)	(1.17)	(1.08)	(1.09)	(0.83)
School×Wave FE	Yes	Yes	Yes	Yes	Yes
Year×Private×Wave FE	Yes	Yes	Yes	Yes	Yes
F-test for $\theta_0=\theta_1=\theta_2=0$	0.0000	0.0000	0.2457	0.0000	0.8121
Observations	8,367	8,356	8,344	8,381	8,379
Pre-reform mean	10.28	8.92	9.59	12.27	2.57

Note: The table shows effects of the Boost for Mathematics on teachers' self-reported peer collaboration activities. All models include school-by-wave fixed effects and time-by-wave fixed effects that vary by municipal and voucher schools. The outcome variable indicated in the column heading is the answer to the survey question: "How often do you, together with another mathematics teacher; (1) plan teaching, (2) follow up on teaching, (3) follow up students' knowledge, (4) discuss instructional practices, or (5) visit each other's lessons to exchange experiences?". Answers are reported as frequency per term. The table reports the p-value of the F-test for the hypothesis that all model parameters are zero. Cluster-adjusted standard errors at the school level are in parentheses and */**/*** refers to statistical significance at the 10/5/1 percent level.

Table C4. Effects of the Boost for Mathematics on aggregated teacher peer collaboration activities (frequency per term)

Outcome:	Aggregated teacher peer collaboration activities
Implementation year	11.24***
	(1.69)
1 year after implementation	3.12
	(2.61)
2 years after implementation	6.38*
	(3.74)
School×Wave FE	Yes
Year×Private×Wave FE	Yes
F-test for $\theta_0=\theta_1=\theta_2=0$	0.0000
Observations	8,285
Pre-reform mean	41.04

Note: The table shows effects of the Boost for Mathematics on aggregated teacher peer collaboration activities. The regression model includes school-by-wave fixed effects and time-by-wave fixed effects that vary by municipal and voucher schools. The outcome variable is the sum of the self-reported answers to the survey question: "How often do you, together with another mathematics teacher; (1) plan teaching, (2) follow up on teaching, (3) follow up students' knowledge, (4) discuss instructional practices?". Answers are reported as frequency per term. The table reports the p-value of the F-test for the hypothesis that all model parameters are zero. Cluster-adjusted standard errors at the school level are in parentheses and */**/*** refers to statistical significance at the 10/5/1 percent level.

Table C5. School-level correlation between changes in teacher peer collaboration and classroom practices

Outcome:	Effect on practices
Effect on collaboration	0.3472**
	(0.1667)
Number of schools	269

Note: The table shows the results from an analysis where the school-level effect of the Boost for Mathematics for classroom practices is regressed on the school-level effect for teacher collaboration. The school-level effects of the program are estimated by taking the change in classroom practices or teacher collaboration, respectively, before and after the implementation. The regression is weighted by 1/(standard error for the estimated effects on teacher collaboration)^2. The measure of classroom practices is calculated as (a) + (b) - (c) with respect to the following survey questions: "In a typical week, what percentage of the lesson time in mathematics do students spend on each of the following activities; (a) discussing problem-solving strategies together with the teacher, (b) other activities, (c) working problems on their own or in group?". The measure of teacher peer collaboration is calculated as the average of the answers to the survey question: "How often do you, together with another mathematics teacher; (a) plan teaching, (b) follow up on teaching, (c) follow up students' knowledge, (d) discuss instructional practices?", where answers are reported as frequency per term.

Table C6. Effects of the Boost for Mathematics on teachers' classroom practices (share of lecture time)

Column:	(1)	(2)	(3)	(4)	(5)
Outcome:	Teacher lectures	Teacher and students discuss	Students solve problems	Students take tests	Other activities
Implementation year	0.04	1.65***	-2.95***	-0.26	1.53***
	(0.46)	(0.60)	(0.76)	(0.24)	(0.51)
1 year after implementation	-0.07	2.79***	-3.84***	-0.17	1.28
	(0.67)	(0.99)	(1.18)	(0.40)	(0.84)
2 years after implementation	0.47	1.86	-3.35**	-0.77	1.78
	(1.02)	(1.44)	(1.59)	(0.56)	(1.09)
School×Wave FE	Yes	Yes	Yes	Yes	Yes
Year×Private×Wave FE	Yes	Yes	Yes	Yes	Yes
F-test for $\theta_0=\theta_1=\theta_2=0$	0.9381	0.0214	0.0008	0.4043	0.0277
Observations	7,816	7,816	7,816	7,816	7,816
Pre-reform mean	18.23	18.81	50.03	5.42	7.51

Note: The table shows effects of the Boost for Mathematics on teachers' self-reported classroom practices. All models include school-by-wave fixed effects and time-by-wave fixed effects that vary by municipal and voucher schools. The outcome variable indicated in the column heading is the answer to the survey question: "In a typical week, what percentage of the lesson time in mathematics do students spend on each of the following activities; (1) listening to lecture-style presentations, (2) discussing problem-solving strategies together with the teacher, (3) working problems on their own or in group, (4) taking tests or quizzes, or (5) other student activities?" Answers are reported as percent of time. The table reports the p-value of the F-test for the hypothesis that all model parameters are zero. Cluster-adjusted standard errors at the school level are in parentheses and */**/*** refers to statistical significance at the 10/5/1 percent level.

Table C7. Effects of the Boost for Mathematics on teachers' training activities (hours per school year) by implementation wave

Column:	(1)	(2)	(3)	(4)
Outcome:	Mathematics	Didactics	Collaboration	Assessment
	2013 wave			
Implementation year	14.71***	22.62***	12.99***	2.45
	(1.43)	(1.72)	(1.85)	(1.84)
1 year after implementation	3.97**	4.00**	2.84	-2.71
	(1.56)	(1.86)	(2.06)	(1.74)
2 years after implementation	1.58	1.08	1.86	-2.42
	(1.53)	(1.87)	(1.94)	(1.77)
Observations	3,222	3,222	3,222	3,222
Pre-reform mean	4.58	6.05	13.45	11.43
	2014 wave			
Implementation year	13.44***	22.81***	10.74***	2.37
	(1.46)	(1.64)	(1.78)	(1.44)
1 year after implementation	1.74	5.53***	0.50	-2.04
	(1.14)	(1.59)	(1.99)	(1.59)
Observations	2,955	2,955	2,955	2,955
Pre-reform mean	4.04	5.17	12.87	10.51
	2015 wave			
Implementation year	15.83***	24.47***	13.93***	4.88***
	(1.60)	(1.69)	(2.13)	(1.64)
Observations	2,196	2,196	2,196	2,196
Pre-reform mean	4.33	5.30	13.28	10.32
School×Wave FE	Yes	Yes	Yes	Yes
Year×Private×Wave FE	Yes	Yes	Yes	Yes
P-value for H_0: $\theta_{0,2013}=\theta_{0,2014}=\theta_{0,2015}$	0.6286	0.6920	0.5336	0.4673
P-value for H_0: $\theta_{0,2013}=\theta_{0,2014}$ and $\theta_{1,2013}=\theta_{1,2014}$	0.2928	0.6701	0.3810	0.8945

Note: The table shows effects of the Boost for Mathematics on teachers' self-reported training activities. All models include school-by-wave fixed effects and time-by-wave fixed effects that vary by municipal and voucher schools. The outcome variable indicated in the column heading is the answer to the survey question: "This academic year, how many hours have you participated in in-service training or other activities that involved; (1) subject knowledge in mathematics, (2) didactics of mathematics, (3) peer collaboration, (4) student assessment?". Answers are reported as hours per school year. The table reports the p-value of the F-test for the hypothesis that all model parameters are zero. Cluster-adjusted standard errors at the school level are in parentheses and */**/*** refers to statistical significance at the 10/5/1 percent level.

Table C8. Effects of the Boost for Mathematics on teacher peer collaboration activities (frequency per term) by implementation wave

Column:	(1)	(2)	(3)	(4)
Outcome:	Plan teaching	Follow-up teaching	Assess students	Discuss didactics
	2013 wave			
Implementation year	4.52***	3.48***	1.58*	5.25***
	(1.02)	(1.01)	(0.94)	(0.93)
1 year after implementation	1.11	1.12	1.24	2.38**
	(1.14)	(1.02)	(1.03)	(1.08)
2 years after implementation	1.67	1.55	1.59	1.76
	(1.20)	(1.17)	(1.08)	(1.09)
Observations	3,221	3,215	3,210	3,224
Pre-reform mean	10.31	8.76	9.54	12.11
	2014 wave			
Implementation year	2.33**	1.46	-0.34	3.45***
	(1.02)	(0.95)	(0.87)	(0.79)
1 year after implementation	0.81	-0.03	-0.49	-0.38
	(1.35)	(1.27)	(1.09)	(1.01)
Observations	2,955	2,952	2,946	2,961
Pre-reform mean	10.04	8.74	9.39	12.11
	2015 wave			
Implementation year	2.65*	3.15**	1.19	4.20***
	(1.43)	(1.49)	(1.19)	(1.40)
Observations	2,191	2,189	2,188	2,196
Pre-reform mean	10.56	9.40	9.93	12.72
School×Wave FE	Yes	Yes	Yes	Yes
Year×Private×Wave FE	Yes	Yes	Yes	Yes
P-value for H_0: $\theta_{0,2013}=\theta_{0,2014}=\theta_{0,2015}$	0.3360	0.3847	0.3568	0.3392
P-value for H_0: $\theta_{0,2013}=\theta_{0,2014}$ and $\theta_{1,2013}=\theta_{1,2014}$	0.3807	0.2539	0.1485	0.0458

Note: The table shows effects of the Boost for Mathematics on teachers' self-reported peer collaboration activities. All models include school-by-wave fixed effects and time-by-wave fixed effects that vary by municipal and voucher schools. The outcome variable indicated in the column heading is the answer to the survey question: "How often do you, together with another mathematics teacher; (1) plan teaching, (2) follow up on teaching, (3) follow up students' knowledge, (4) discuss instructional practices?". Answers are reported as frequency per term. The table reports the p-value of the F-test for the hypothesis that all model parameters are zero. Cluster-adjusted standard errors at the school level are in parentheses and */**/*** refers to statistical significance at the 10/5/1 percent level.

Table C9. Effects of the Boost for Mathematics on teachers' classroom practices (share of lecture time) by wave

Column:	(1)	(2)	(3)	(4)
Outcome:	Teacher lectures	Teacher and students discuss	Students solve problems	Students take tests
	2013 wave			
Implementation year	1.34	0.72	-3.81**	-0.77
	(1.02)	(1.27)	(1.59)	(0.47)
1 year after implementation	1.07	3.30**	-5.25***	-0.64
	(0.92)	(1.37)	(1.70)	(0.52)
2 years after implementation	0.47	1.86	-3.35**	-0.77
	(1.02)	(1.45)	(1.59)	(0.56)
Observations	3,009	3,009	3,009	3,009
Pre-reform mean	18.36	18.64	50.18	5.58
	2014 wave			
Implementation year	-0.11	3.29**	-2.63*	0.05
	(0.86)	(1.41)	(1.52)	(0.47)
1 year after implementation	-1.46	2.17	-2.12	0.41
	(1.00)	(1.42)	(1.70)	(0.50)
Observations	2,761	2,761	2,761	2,761
Pre-reform mean	18.14	18.79	49.89	5.43
	2015 wave			
Implementation year	-2.51**	0.68	-1.67	0.26
	(1.14)	(1.37)	(1.81)	(0.54)
Observations	2,046	2,046	2,046	2,046
Pre-reform mean	18.14	19.08	50.02	5.19
School×Wave FE	Yes	Yes	Yes	Yes
Year×Private×Wave FE	Yes	Yes	Yes	Yes
P-value for H_0: $\theta_{0,2013}=\theta_{0,2014}=\theta_{0,2015}$	0.0351	0.4221	0.7141	0.2880
P-value for H_0: $\theta_{0,2013}=\theta_{0,2014}$ and $\theta_{1,2013}=\theta_{1,2014}$	0.1345	0.6578	0.3156	0.1206

Note: The table shows effects of the Boost for Mathematics on teachers' self-reported classroom practices. All models include school-by-wave fixed effects and time-by-wave fixed effects that vary by municipal and voucher schools. The outcome variable indicated in the column heading is the answer to the survey question: "In a typical week, what percentage of the lesson time in mathematics do students spend on each of the following activities; (1) listening to lecture-style presentations, (2) discussing problem-solving strategies together with the teacher, (3) working problems on their own or in group, (4) taking tests or quizzes?" Answers are reported as percent of time. The table reports the p-value of the F-test for the hypothesis that all model parameters are zero. Cluster-adjusted standard errors at the school level are in parentheses and */**/*** refers to statistical significance at the 10/5/1 percent level.

Table C10. Effects of the Boost for Mathematics on teachers' sources of inspiration for improving their instruction

Column:	(1)	(2)	(3)	(4)	(5)
Outcome:	School management	Colleagues	Educational web-platforms	Literature	Seminars and conferences
Implementation year	0.43*	3.71***	0.75	2.00***	2.01***
	(0.23)	(0.47)	(0.47)	(0.44)	(0.28)
1 year after implementation	-0.03	1.95**	0.63	0.73	0.01
	(0.38)	(0.78)	(0.76)	(0.67)	(0.38)
2 years after implementation	0.25	1.69	1.68	0.13	0.45
	(0.49)	(1.19)	(1.10)	(0.98)	(0.50)
School×Wave FE	Yes	Yes	Yes	Yes	Yes
Year×Private×Wave FE	Yes	Yes	Yes	Yes	Yes
F-test for $\theta_0=\theta_1=\theta_2=0$	0.1599	0.0000	0.3640	0.0000	0.0000
Observations	8,086	8,312	8,247	8,292	8,257
Pre-reform mean	1.66	13.64	9.11	7.83	2.38

Note: The table shows effects of the Boost for Mathematics on teachers' self-reported sources of inspiration for improving their instruction. All models include school-by-wave fixed effects and time-by-wave fixed effects that vary by municipal and voucher schools. The outcome variable indicated in the column heading is the answer to the survey question: "How often do you get inspiration and knowledge to improve your instruction from; (1) the school management, (2), colleagues, (3) educational web-platforms, (4) literature (e.g., books and research papers), or (5) seminars and conferences?" Answers are reported as: "At least once a week" (25 times per semester); "At least once a month" (12); "At least once per semester" (3); "More rarely/never" (0). The table reports the p-value of the F-test for the hypothesis that all model parameters are zero. Cluster-adjusted standard errors at the school level are in parentheses and */**/*** refers to statistical significance at the 10/5/1 percent level.

Table C11. Effects of the Boost for Mathematics on teachers' self-assessment of their knowledge and competences

Column:	(1)	(2)	(3)
Outcome:	Subject knowledge in Mathematics	Mathematics didactics	Assessing the results of teaching
Implementation year	0.11**	0.18***	0.20***
	(0.05)	(0.05)	(0.04)
1 year after implementation	0.10	0.01	0.12
	(0.07)	(0.08)	(0.07)
2 years after implementation	0.21*	0.08	0.21**
	(0.11)	(0.11)	(0.09)
School×Wave FE	Yes	Yes	Yes
Year×Private×Wave FE	Yes	Yes	Yes
F-test for $\theta_0=\theta_1=\theta_2=0$	0.0913	0.0001	0.0001
Observations	8,390	8,371	8,362

Note: The table shows effects of the Boost for Mathematics on teachers' self-assessment of their knowledge and competences. All models include school-by-wave fixed effects and time-by-wave fixed effects that vary by municipal and voucher schools. The outcome variable indicated in the column heading is the answer to the survey question: "To what extent do you think you have sufficient knowledge and competence in; (1) mathematics, (2) methodology and didactics of mathematics, or (3) following up the results of your mathematics teaching?" Answers are reported as: "To a very high degree" (5); "To a high degree" (4); "To neither a high nor a low degree" (3); "To a low degree" (1); "To a very low degree" (1). The outcome variable has been standardized. The table reports the p-value of the F-test for the hypothesis that all model parameters are zero. Cluster-adjusted standard errors at the school level are in parentheses and */**/*** refers to statistical significance at the 10/5/1 percent level.

Table C12. Effects of the Boost for Mathematics on teachers' opinion of their school

Column:	(1)	(2)	(3)
Outcome:	Principals' pedagogical leadership	Colleagues' subject knowledge in Mathematics	Colleagues' knowledge in didactics of Mathematics
Implementation year	-0.00	0.11*	0.20***
	(0.06)	(0.06)	(0.06)
1 year after implementation	-0.10	0.03	0.18**
	(0.12)	(0.09)	(0.09)
2 years after implementation	-0.03	0.14	0.22*
	(0.16)	(0.15)	(0.13)
School×Wave FE	Yes	Yes	Yes
Year×Private×Wave FE	Yes	Yes	Yes
F-test for $\theta_0=\theta_1=\theta_2=0$	0.6021	0.1474	0.0046
Observations	8,073	8,020	7,891

Note: The table shows effects of the Boost for Mathematics on teachers' self-reported opinion of their school. All models include school-by-wave fixed effects and time-by-wave fixed effects that vary by municipal and voucher schools. The outcome variable indicated in the column heading is the answer to the survey question: "How do you think the following is at your school; (1) the principal's pedagogical leadership, (2) the mathematics teachers' subject knowledge in mathematics, and (3) the mathematics teachers' knowledge of methodology and didactics in mathematics?" Answers are reported as: "Very good" (5); "Good" (4); "Neither good nor bad" (3); "Bad" (2); "Very bad" (1). The outcome variable has been standardized. The table reports the p-value of the F-test for the hypothesis that all model parameters are zero. Cluster-adjusted standard errors at the school level are in parentheses and */**/*** refers to statistical significance at the 10/5/1 percent level.

Table C13. Effects of the Boost for Mathematics on teachers' training activities (hours per school year) for schools with a student population below and above the median in predicted test scores

Column:	(1)	(2)	(3)	(4)
Outcome:	Mathematics	Didactics	Collaboration	Assessment
	Weak schools			
Implementation year	14.06***	21.78***	12.47***	3.02**
	(1.07)	(1.29)	(1.17)	(1.25)
1 year after implementation	3.83***	4.91***	3.15*	0.37
	(1.15)	(1.76)	(1.74)	(1.88)
2 years after implementation	1.40	-0.78	1.60	0.01
	(1.62)	(2.42)	(2.55)	(2.55)
F-test for $\theta_0=\theta_1=\theta_2=0$	0.0000	0.0000	0.0000	0.0086
Observations	3,890	3,890	3,890	3,890
Pre-reform mean	3.84	5.08	11.59	9.54
	Strong schools			
Implementation year	14.61***	23.55***	12.61***	2.51**
	(1.22)	(1.27)	(1.17)	(1.13)
1 year after implementation	2.40	5.31***	2.36	-3.94***
	(1.84)	(2.03)	(1.91)	(1.34)
2 years after implementation	1.93	2.32	3.42	-4.06*
	(2.73)	(3.01)	(2.78)	(2.42)
F-test for $\theta_0=\theta_1=\theta_2=0$	0.0000	0.0000	0.0000	0.0000
Observations	3,879	3,879	3,879	3,879
Pre-reform mean	4.72	6.06	14.38	12.07
School×Wave FE	Yes	Yes	Yes	Yes
Year×Private×Wave FE	Yes	Yes	Yes	Yes
P-value for $H_0: \theta^{weak}=\theta^{strong}$	0.9534	0.3699	0.8729	0.2371

Note: The table shows effects of the Boost for Mathematics on teachers' self-reported training activities. All models include school-by-wave fixed effects and time-by-wave fixed effects that vary by municipal and voucher schools. The outcome variable indicated in the column heading is the answer to the survey question: "This academic year, how many hours have you participated in in-service training or other activities that involved; (1) subject knowledge in mathematics, (2) didactics of mathematics, (3) peer collaboration, (4) student assessment?". Answers are reported as hours per school year. The table reports the p-value of the F-test for the hypothesis that all model parameters are zero. Cluster-adjusted standard errors at the school level are in parentheses and */**/*** refers to statistical significance at the 10/5/1 percent level.

Table C14. Effects of the Boost for Mathematics on teacher peer collaboration activities (frequency per term) for schools with a student population below and above the median in predicted test scores

Column:	(1)	(2)	(3)	(4)
Outcome:	Plan teaching	Follow-up teaching	Assess students	Discuss didactics
	Weak schools			
Implementation year	4.04***	3.09***	1.35*	5.28***
	(0.88)	(0.84)	(0.73)	(0.72)
1 year after implementation	1.88	1.89*	1.00	1.50
	(1.32)	(1.10)	(0.95)	(1.07)
2 years after implementation	3.52*	1.77	1.67	1.68
	(1.79)	(1.61)	(1.55)	(1.64)
F-test for $\theta_0=\theta_1=\theta_2=0$	0.0000	0.0016	0.3323	0.0000
Observations	3,890	3,882	3,874	3,892
Pre-reform mean	9.38	8.37	9.39	11.64
	Strong schools			
Implementation year	2.66***	2.23***	0.49	4.15***
	(0.89)	(0.83)	(0.73)	(0.80)
1 year after implementation	0.22	-0.97	-0.24	1.69
	(1.39)	(1.35)	(1.24)	(1.10)
2 years after implementation	0.14	1.42	1.56	2.77*
	(1.79)	(1.82)	(1.67)	(1.52)
F-test for $\theta_0=\theta_1=\theta_2=0$	0.0012	0.0004	0.3806	0.0000
Observations	3,873	3,875	3,868	3,884
Pre-reform mean	10.70	9.03	9.47	12.86
School×Wave FE	Yes	Yes	Yes	Yes
Year×Private×Wave FE	Yes	Yes	Yes	Yes
P-value for $H_0: \theta^{weak}=\theta^{strong}$	0.2385	0.3092	0.4985	0.8679

Note: The table shows effects of the Boost for Mathematics on teachers' self-reported peer collaboration activities. All models include school-by-wave fixed effects and time-by-wave fixed effects that vary by municipal and voucher schools. The outcome variable indicated in the column heading is the answer to the survey question: "How often do you, together with another mathematics teacher; (1) plan teaching, (2) follow up on teaching, (3) follow up students' knowledge, (4) discuss instructional practices?". Answers are reported as frequency per term. The table reports the p-value of the F-test for the hypothesis that all model parameters are zero. Cluster-adjusted standard errors at the school level are in parentheses and */**/*** refers to statistical significance at the 10/5/1 percent level.

Table C15. Effects of the Boost for Mathematics on teachers' classroom practices (share of lecture time) for schools with a student population below and above the median in predicted test scores

Column:	(1)	(2)	(3)	(4)
Outcome:	Teacher lectures	Teacher and students discuss	Students solve problems	Students take tests
	Weak schools			
Implementation year	0.61	2.19**	-4.54***	-0.55
	(0.71)	(1.06)	(1.19)	(0.40)
1 year after implementation	0.54	3.21*	-4.97***	0.28
	(1.02)	(1.68)	(1.87)	(0.67)
2 years after implementation	0.29	2.72	-4.87**	-0.42
	(1.60)	(2.15)	(2.46)	(0.84)
F-test for $\theta_0=\theta_1=\theta_2=0$	0.7856	0.1751	0.0014	0.0848
Observations	3,620	3,620	3,620	3,620
Pre-reform mean	17.63	19.12	49.92	5.74
	Strong schools			
Implementation year	-0.40	1.17	-1.68*	-0.15
	(0.63)	(0.79)	(1.00)	(0.34)
1 year after implementation	0.04	2.56*	-3.22*	-0.95*
	(0.97)	(1.34)	(1.70)	(0.54)
2 years after implementation	1.07	-0.62	-0.57	-1.20
	(1.36)	(2.10)	(2.08)	(0.80)
F-test for $\theta_0=\theta_1=\theta_2=0$	0.5697	0.1559	0.1205	0.1464
Observations	3,645	3,645	3,645	3,645
Pre-reform mean	18.94	18.55	50.13	5.05
School×Wave FE	Yes	Yes	Yes	Yes
Year×Private×Wave FE	Yes	Yes	Yes	Yes
P-value for $H_0: \theta^{weak}=\theta^{strong}$	0.6397	0.4280	0.1504	0.5944

Note: The table shows effects of the Boost for Mathematics on teachers' self-reported classroom practices. All models include school-by-wave fixed effects and time-by-wave fixed effects that vary by municipal and voucher schools. The outcome variable indicated in the column heading is the answer to the survey question: "In a typical week, what percentage of the lesson time in mathematics do students spend on each of the following activities; (1) listening to lecture-style presentations, (2) discussing problem-solving strategies together with the teacher, (3) working problems on their own or in group, (4) taking tests or quizzes?" Answers are reported as percent of time. The table reports the p-value of the F-test for the hypothesis that all model parameters are zero. Cluster-adjusted standard errors at the school level are in parentheses and */**/*** refers to statistical significance at the 10/5/1 percent level.

Table C16. Effects of the Boost for Mathematics on teacher peer collaboration activities (frequency per term) for teachers with below and above median experience

Column:	(1)	(2)	(3)	(4)
Outcome:	Plan teaching	Follow-up teaching	Assess students	Discuss didactics
	Below median experience			
Implementation year	2.73***	1.97***	0.19	3.78***
	(0.74)	(0.73)	(0.70)	(0.75)
1 year after implementation	-0.47	-0.88	-0.98	-0.33
	(1.20)	(1.10)	(1.04)	(1.13)
2 years after implementation	1.44	-0.50	0.07	-0.48
	(1.56)	(1.61)	(1.50)	(1.56)
F-test for θ_0=θ_1=θ_2=0	0.0000	0.0005	0.4161	0.0000
Observations	4,366	4,365	4,344	4,371
Pre-reform mean	9.51	8.02	8.91	11.93
	Above median experience			
Implementation year	3.65***	2.87***	0.89	4.44***
	(0.76)	(0.70)	(0.62)	(0.69)
1 year after implementation	2.12*	1.53	1.07	1.85*
	(1.14)	(1.07)	(1.03)	(0.95)
2 years after implementation	1.68	3.23**	2.94**	4.06***
	(1.65)	(1.61)	(1.45)	(1.49)
F-test for θ_0=θ_1=θ_2=0	0.0000	0.0005	0.2083	0.0000
Observations	3,835	3,816	3,823	3,838
Pre-reform mean	11.18	9.96	10.38	12.68
School×Wave FE	Yes	Yes	Yes	Yes
Year×Private×Wave FE	Yes	Yes	Yes	Yes
P-value for H_0: $\theta^{below}=\theta^{above}$	0.2623	0.0911	0.1717	0.1065

Note: The table shows effects of the Boost for Mathematics on teachers' self-reported peer collaboration activities. All models include school-by-wave fixed effects and time-by-wave fixed effects that vary by municipal and voucher schools. The outcome variable indicated in the column heading is the answer to the survey question: "How often do you, together with another mathematics teacher; (1) plan teaching, (2) follow up on teaching, (3) follow up students' knowledge, (4) discuss instructional practices?". Answers are reported as frequency per term. The table reports the p-value of the F-test for the hypothesis that all model parameters are zero. Cluster-adjusted standard errors at the school level are in parentheses and */**/*** refers to statistical significance at the 10/5/1 percent level.

Table C17. Effects of the Boost for Mathematics on teacher peer collaboration activities (frequency per term) for teachers below and above median mathematics education

Column:	(1)	(2)	(3)	(4)
Outcome:	Plan teaching	Follow-up teaching	Assess students	Discuss didactics
	Below median mathematics education			
Implementation year	4.09***	3.32***	0.42	4.60***
	(0.79)	(0.79)	(0.72)	(0.81)
1 year after implementation	-0.14	-0.47	-0.95	0.54
	(1.36)	(1.31)	(1.13)	(1.24)
2 years after implementation	2.26	1.43	0.84	2.22
	(1.82)	(1.77)	(1.67)	(1.93)
F-test for $\theta_0=\theta_1=\theta_2=0$	0.0000	0.0000	0.3424	0.0000
Observations	3,902	3,897	3,884	3,911
Pre-reform mean	9.50	8.56	9.16	12.20
	Above median mathematics education			
Implementation year	3.50***	2.78***	1.24*	4.76***
	(0.79)	(0.76)	(0.71)	(0.65)
1 year after implementation	2.64**	1.97*	1.30	2.13**
	(1.15)	(1.11)	(1.10)	(1.00)
2 years after implementation	2.49	2.45	2.19	2.75**
	(1.55)	(1.70)	(1.67)	(1.28)
F-test for $\theta_0=\theta_1=\theta_2=0$	0.0001	0.0013	0.3590	0.0000
Observations	4,352	4,349	4,353	4,360
Pre-reform mean	11.16	9.33	10.09	12.37
School×Wave FE	Yes	Yes	Yes	Yes
Year×Private×Wave FE	Yes	Yes	Yes	Yes
P-value for $H_0: \theta^{below}=\theta^{above}$	0.5885	0.5722	0.2792	0.5503

Note: The table shows effects of the Boost for Mathematics on teachers' self-reported peer collaboration activities. All models include school-by-wave fixed effects and time-by-wave fixed effects that vary by municipal and voucher schools. The outcome variable indicated in the column heading is the answer to the survey question: "How often do you, together with another mathematics teacher; (1) plan teaching, (2) follow up on teaching, (3) follow up students' knowledge, (4) discuss instructional practices?". Answers are reported as frequency per term. The table reports the p-value of the F-test for the hypothesis that all model parameters are zero. Cluster-adjusted standard errors at the school level are in parentheses and */**/*** refers to statistical significance at the 10/5/1 percent level.

Table C18. Effects of the Boost for Mathematics on teachers' classroom practices (share of lecture time) for teachers with below and above median experience

Column:	(1)	(2)	(3)	(4)
Outcome:	Teacher lectures	Teacher and students discuss	Students solve problems	Students take tests
	Below median experience			
Implementation year	0.18	0.79	-3.21**	0.37
	(0.77)	(0.90)	(1.27)	(0.40)
1 year after implementation	0.25	2.54	-5.02**	0.48
	(1.10)	(1.61)	(2.03)	(0.59)
2 years after implementation	-0.06	1.45	-3.59	-0.23
	(1.80)	(2.20)	(2.65)	(0.81)
F-test for $\theta_0=\theta_1=\theta_2=0$	0.9841	0.3938	0.0477	0.2854
Observations	4,060	4,060	4,060	4,060
Pre-reform mean	18.07	17.84	50.73	5.53
	Above median experience			
Implementation year	0.03	1.95**	-2.22**	-0.66**
	(0.57)	(0.80)	(0.98)	(0.33)
1 year after implementation	-0.44	2.56*	-1.44	-0.66
	(0.90)	(1.34)	(1.47)	(0.51)
2 years after implementation	0.38	2.34	-2.15	-1.12
	(1.36)	(1.82)	(2.41)	(0.79)
F-test for $\theta_0=\theta_1=\theta_2=0$	0.8574	0.0949	0.1451	0.2556
Observations	3,578	3,578	3,578	3,578
Pre-reform mean	18.33	19.91	49.35	5.28
School×Wave FE	Yes	Yes	Yes	Yes
Year×Private×Wave FE	Yes	Yes	Yes	Yes
P-value for $H_0: \theta^{below}=\theta^{above}$	0.8476	0.6794	0.3309	0.1200

Note: The table shows effects of the Boost for Mathematics on teachers' self-reported peer collaboration activities. All models include school-by-wave fixed effects and time-by-wave fixed effects that vary by municipal and voucher schools. The outcome variable indicated in the column heading is the answer to the survey question: "How often do you, together with another mathematics teacher; (1) plan teaching, (2) follow up on teaching, (3) follow up students' knowledge, (4) discuss instructional practices?". Answers are reported as frequency per term. The table reports the p-value of the F-test for the hypothesis that all model parameters are zero. Cluster-adjusted standard errors at the school level are in parentheses and */**/*** refers to statistical significance at the 10/5/1 percent level.

Table C19. Effects of the Boost for Mathematics on teachers' classroom practices (share of lecture time) for teachers below and above median mathematics education

Column:	(1)	(2)	(3)	(4)
Outcome:	Teacher lectures	Teacher and students discuss	Students solve problems	Students take tests
	Below median mathematics education			
Implementation year	0.32	2.42***	-3.61***	-0.27
	(0.71)	(0.92)	(1.14)	(0.40)
1 year after implementation	1.55	1.44	-2.85	-0.05
	(1.22)	(1.31)	(1.80)	(0.59)
2 years after implementation	-0.85	1.47	-0.12	-0.64
	(1.71)	(2.15)	(2.81)	(0.97)
F-test for $\theta_0=\theta_1=\theta_2=0$	0.0986	0.0504	0.0018	0.7902
Observations	3,689	3,689	3,689	3,689
Pre-reform mean	17.32	19.14	51.10	5.14
	Above median mathematics education			
Implementation year	0.11	0.61	-1.86*	-0.51
	(0.67)	(1.00)	(1.12)	(0.32)
1 year after implementation	-1.10	3.00*	-3.13*	-0.43
	(0.94)	(1.59)	(1.65)	(0.49)
2 years after implementation	1.15	1.89	-4.34**	-1.07*
	(1.56)	(1.96)	(2.05)	(0.65)
F-test for $\theta_0=\theta_1=\theta_2=0$	0.2160	0.1806	0.1740	0.3274
Observations	4,008	4,008	4,008	4,008
Pre-reform mean	19.30	18.42	48.76	5.76
School×Wave FE	Yes	Yes	Yes	Yes
Year×Private×Wave FE	Yes	Yes	Yes	Yes
P-value for $H_0: \theta^{below}=\theta^{above}$	0.6046	0.8892	0.9876	0.5809

Note: The table shows effects of the Boost for Mathematics on teachers' self-reported peer collaboration activities. All models include school-by-wave fixed effects and time-by-wave fixed effects that vary by municipal and voucher schools. The outcome variable indicated in the column heading is the answer to the survey question: "How often do you, together with another mathematics teacher; (1) plan teaching, (2) follow up on teaching, (3) follow up students' knowledge, (4) discuss instructional practices?". Answers are reported as frequency per term. The table reports the p-value of the F-test for the hypothesis that all model parameters are zero. Cluster-adjusted standard errors at the school level are in parentheses and */**/*** refers to statistical significance at the 10/5/1 percent level.

D. Reliability of test scores

In Swedish schools, mathematics teachers grade their own students' national exams. A relevant question is therefore whether the estimated effects of the Boost for Mathematics on test scores reflect changes in teachers' grading standards rather than improved student performance. Even though the Swedish National Agency for Education provides detailed guidelines on how to assess different answers, and promotes co-grading, it is still possible that participating teachers adopt less (or more) stringent grading standards.[1] In this section, we provide three pieces of evidence suggesting that the estimated effects are likely to reflect improved student performance rather than changes in teachers' grading standards.

First, there is little room for teachers' subjective judgement of students' answers to questions that can be characterized as being either "right" or "wrong", as is often the case in mathematics. This is confirmed by the re-assessments of national exams conducted by the Swedish Schools Inspectorate for a sample of schools every year (see e.g., Skolinspektionen 2021). Teachers are often found to be more lenient when judging their own students than are the external graders, but the magnitudes differ considerably across subjects. In Swedish, the deviation in test scores between the teacher and the external examiner is on average more than 20 percent of a standard deviation (of the externally graded test score). The corresponding number for the national exams in mathematics is about 5 percent of a standard deviation (Skolinspektionen 2012). Thus, the teachers' judgement of their own students' mathematics performance does not differ much from that of external examiners.

Second, to further investigate the subjectiveness of teachers' assessment, we make use of data from TIMSS, which is an international assessment of student performance in mathematics and science in grades 4 and 8, conducted by the

[1] To the extent that the program helps teachers to make more reliable (less noisy) assessments of student performance, this would not bias the estimates (but rather make them more precise).

International Association for the Evaluation of Educational Achievement (IEA). We have access to the TIMSS 2015 survey for Sweden, matched to the students' national exams in grades 3, 6 and 9 (Skolverket, 2015). This enables us to compare the effect of teachers (intraclass correlations) for student performance in mathematics on the national tests (internally graded) and on the TIMSS test (externally graded). We find that the teacher effects are of the same order of magnitude for both tests; 0.265 (0.014) for the national tests and 0.249 (0.013) for TIMSS, which, again, suggests that there is little room for teachers' subjective grading in mathematics.

Third, as a final check for any impact of the Boost for Mathematics on teachers' grading standards, we exploit information on program participation obtained from the Swedish version of the TIMSS 2015 school questionnaire (Skolverket, 2015). The schools were asked to state the share of mathematics teachers who participated in the Boost for mathematics (in the 2013/14 or in the 2014/15 school years). Similar to our main analysis, we define schools where at least half of the teachers participate in the program as treated, and schools with no participating teachers as untreated. We can, thus, compare the difference in student performance in mathematics between participating and non-participating schools using both the internally graded national exams (in grades 3 and 9) and the externally graded TIMSS test (in grades 4 and 8).[2]

Appendix Table D1, column 1, shows that students in schools participating in the Boost for Mathematics perform on average about 0.05 SD better on the national tests in mathematics than students in schools that do not participate. However, the difference is not significant. In column 2, we attempt to adjust for some of the selection to the program by adding pre-determined student characteristics, which reduces the differences between schools slightly. In the last two columns of Appendix Table D1, we repeat the same exercise using the

[2] Since TIMSS 2015 is a cross-sectional data set we are unable to control for fixed differences between schools, and the difference in performance between treated and untreated schools may therefore not be given a causal interpretation.

externally graded TIMSS test. Column 3 reveals that students in treated schools score on average about 0.05 SD higher than other students on the TIMSS test (not significant). Again, adding pre-determined student characteristics reduces the estimates somewhat.[3] Thus, the estimated difference in student performance between schools participating in the Boost for Mathematics, and schools not participating, is very similar if we use the internally graded national exams or the externally graded TIMSS test, which indicates that the program had very minor, if any, effects on teachers' grading standards in mathematics.

Table D1. Descriptive differences in student performance in mathematics between schools participating in the Boost for Mathematics and schools that do not, using data on national tests and the TIMSS test

Column:	(1)	(2)	(3)	(4)
Outcome:	Test scores	Test scores	TIMSS test	TIMSS test
Grades:	3 and 9	3 and 9	4 and 8	4 and 8
All years pooled	0.047	0.036	0.047	0.029
	(0.067)	(0.058)	(0.063)	(0.041)
Student controls	No	Yes	No	Yes
Number of students	7,142	7,142	7,581	7,581
Number of schools	270	270	270	270

Note: The table shows differences in student performance in mathematics for schools participating in the Boost for Mathematics and schools that do not participate, using data on national tests (grades 3 and 9) and the TIMSS 2015 test (grades 4 and 8), respectively. The data have been provided by the National Agency of Education. Schools are defined as being treated if at least half of the mathematics teachers in the school participate in the program, and untreated if no teacher participate. All stages have been pooled and the models include a dummy variable for grade level. The student controls consist of dummy variables for month of birth, gender, first- and second-generation immigrant, age at immigration, and mother's and father's highest educational level. The outcome variable studied is indicated in the column heading. Cluster-adjusted standard errors at the school level are in parentheses and */**/*** refers to statistical significance at the 10/5/1 percent level.

[3] This result is consistent with Lindvall et al. (2022) that also use data from TIMSS 2015. They do not find any significant performance differences between students taught by teachers participating in the Boost for Mathematics and other students.

E. Cost and benefit calculations

The evaluation of the Boost for Mathematics captures the short-run effects on mathematics skills. To inform policy about the efficiency of the program, however, it is necessary to also take the costs and long-run benefits of the intervention into account. In this section, we attempt to attach a monetary value to the societal costs and benefits of the program compared to the situation had it not been introduced.

Costs

During the implementation phase of the Boost for Mathematics, participating teachers devote about 60 hours of their time to the learning cycles. The training is required to take place during regular working hours, and is, thus, expected to crowd out other out-of-class teacher activities. We lack time-use data for participating teachers but assume that half of their non-teaching activities – such as preparation, interaction with students and parents, and other types of professional development – are directly (or indirectly) related to students' human capital production, whereas the other half – such as school management, administration, and extracurricular activities – produce other outputs valuable to society.

To the extent that the program infringes on out-of-class activities that matter for skill formation, this will be captured by the estimated test score effects. This is, however, not the case for other types of teacher activities, and we therefore value the lost production of other societal goods by the market price of teacher time. The gross hourly wage (including payroll taxes) for participating teachers is €28.6 (in 2020 prices). Since we assume that half of the 60 training hours crowd out production of other societal goods, we estimate the cost of training at €858 per teacher (€28.6×60×0.50). In all, 23,209 teachers participated in the program in the schools covered by the evaluation, and the total cost for all teachers is, thus, about €19.9 million (€858 ×23,209).

The external tutors are expected to spend 20 percent of their time to prepare and coach teachers, which corresponds to about 400 hour per school year. We take the full opportunity cost of the time the external tutors spend on the program into account, since it is not likely to affect student performance in treated schools. Assuming that tutors have the same wage as the average participating teacher, the cost is estimated at €11,440 per tutor (€28.6×400). There were 1,360 tutors hired in the schools covered by the evaluation, adding up to a cost of about €15.6 million (€11,440×1,360).

The program also involved other costs, such as the training of tutors and principals, setting up the web-portal, administration, etc., amounting to €15.7 million (Skolverket 2016).[4] The grand total cost of the program is, thus, estimated to be about €51.3 million (€19.9 + €15.6 + €15.7 million). In all, 646,276 unique students were exposed to the program at some point (in the schools covered by the evaluation), yielding an average cost per student of about €80 (€51.2 million / 646,276 students).

Benefits

The major benefit of the Boost for mathematics is the value of the students' improved mathematics performance. We translate the short-run learning effects to life-time earning gains using data from the "Evaluation-through-follow-up" (ETF) project (Gothenburg University, 1985). The ETF data includes information on, among other things, mathematics performance and cognitive abilities in grade 6 for a 10 percent sample of cohorts born 1953, 1967, and 1972. The individuals are matched to their family background (Statistics Sweden, 1990; 2019a; 2019c; 2019d) and earnings records for the 1968–2015 period (Statistics Sweden, 1990; 2019b), making it possible to follow the earliest cohort throughout most of their labor market careers. We calculate the present value of life-cycle earnings by discounting the real annual earnings (including payroll taxes)

[4] We assume that the accounting cost corresponds to the value of lost production.

in the period 1968–2015 at 3 percent (in 2020 prices). For ease of interpretation, we divide the life-cycle earnings by the mean (separately by cohort and gender), and the estimates should be interpreted as a percentage change associated with 1 SD better mathematics skills.

Table E1. The life-cycle earnings associated with mathematics skills in grade 6

Column:	(1)	(2)	(3)	(4)	(5)	(6)
Model:	OLS	OLS	Twin FE	IV	IV	Twin FE-IV
			Panel A. All in 1953 cohort			
Mathematics test score (SD)	0.092***	0.082***		0.107***	0.096***	
	(0.005)	(0.005)		(0.007)	(0.007)	
Individual controls	No	Yes		No	Yes	
R^2	0.048	0.083		0.047	0.082	
Observations	8,090	8,090		8,090	8,090	
			Panel B. Twins in 1953–72 cohorts			
Mathematics test score (SD)	0.067***	0.071***	0.065*	0.091***	0.097**	0.109*
	(0.020)	(0.023)	(0.035)	(0.034)	(0.039)	(0.063)
Individual controls	No	Yes	Yes	No	Yes	Yes
R^2	0.039	0.158	0.680	0.035	0.155	0.675
Observations	354	354	354	354	354	354

Note: The table shows the association between the present value of real life-cycle earnings and standardized mathematics test scores in grade 6. The present value of life-cycle earnings has been obtained by discounting real annual earnings in the period 1968–2015 (in 2020 prices) at 3 percent. The life-cycle earnings have been divided by the mean in the population (by cohort and gender), and the estimates should be interpreted as a percentage change associated with 1 SD better mathematics skills. All models control for gender and cohort. The individual controls are dummy variables for month of birth, indicators for first and second generation immigrant, dummy variables for age at immigration, dummy variables for mother's and father's highest level of education, mother's and father's percentile rank mid-age (35–45 years) earnings in levels and squared, and indicator variables for having missing information on mother's or father's earnings. Columns (4)–(6) attempt to adjust for measurement error using the individual's logical-inductive ability in grade 6 as an instrument for mathematics test scores in grade 6. */**/*** refers to statistical significance at the 10/5/1 percent level.

Table E1 shows the life-cycle earnings associated with 1 SD higher mathematics test score in grade 6. All models control for gender and cohort fixed effects.

Panel A shows the estimates for individuals born in 1953, for whom we can observe earnings for ages 16–62. The first column shows that 1 SD better mathematics performance in grade 6 is associated with about 9 percent higher life-cycle earnings. The second column accounts for differences in observed demographic characteristics and family background, which leads to slightly lower estimates. Columns 4 and 5 attempt to adjust for measurement error in the observed mathematics scores by using the individual's logical-inductive ability in grade 6 as an instrument. This increases the estimates slightly, and 1 SD better test scores is associated with about 9 percent higher discounted real life-cycle earnings.

The ETF-data includes a sample of twins which allows us to also account for unobserved family characteristics. Panel B of Table E1 shows the estimates for twins born 1953, 1967, and 1972, for whom we observe parts of their labor market career. Columns 1–2 and 4–5 replicate the models used in Panel A for the twin sample. Column 3 shows that the estimates are substantially reduced when adding twin FE to the model, which indicates that the association between test scores and earnings partly reflects difference in unobserved family background. An alternative explanation, however, is that the potential bias arising from measurement errors in observed test scores is exacerbated when exploiting the within-twin variation. This is supported by the results presented in the last column, where we use logical-inductive ability as an instrument for mathematics test scores in an attempt to adjust for attenuation bias. This leads to an association of about 10 percent but it is rather imprecisely estimated. Thus, unobserved (or observed) family background does not seem to drive much of the correlation between test scores and earnings.

Based on these estimates we assume that the return to 1 SD better mathematics performance over the life-cycle amounts to 9 percent.[5] In our data, the

[5] (Öckert 2021) reviews papers attempting to estimate causal effects of educational attainment on skills and earnings and finds that, on average, one year of schooling improves test scores by

average real gross life-cycle earnings (including employer contributions), discounted at 3 percent to age 16, for men born 1952–53 is about €940,000 (in 2020 prices). We arrive at an estimated benefit of the Boost for Mathematics by first dividing the reduced form effect for different years of exposure (first column of Appendix Table B3) by the share of treated students (first column of Appendix Table B2) and then multiplying by the estimated return to test scores (Appendix Table E1), the discounted life-cycle earnings (discounted back to the age when students are first exposed to the program) and, finally, the number of students. This yields an estimated benefit of about €1,386 million, or about €2,144 per student (€1,386 million/646,276 students). The benefit-to-cost ratio is about 27 (€1,386 million/€51.3 million), meaning that the program generates €27 in savings for every €1 spent.

The calculations suggest that the Boost for Mathematics passes a cost-benefit test. It should be stressed, however, that the estimated societal benefits and costs are uncertain, and the effectiveness of the program may change under alternative assumptions. For instance, we base the benefit calculations only on students who have taken the final exams by year 2019, while the results show that test scores improve also for students who enter school after program implementation. Thus, if we were to extrapolate the effects of the program also for future incoming cohorts, the benefits of the Boost for Mathematics would increase even further.

On the other hand, our calculations may overstate the productivity gains of the program. Some of the estimated return to mathematics skills in Table E1 could reflect sorting of individuals in the education system – along with the corresponding return to schooling – as well as signaling on the labor market. In addition, the program could generate general equilibrium effects on the labor market, which would dampen the productivity gains. However, even if only half

about 0.25 SD and earnings by 2.5 percent. This leads to an earnings-to-skill-effects-ratio of 10 percent.

of the estimated return to skills is due to improved productivity, the benefit-cost-ratio would still be more than 13. Thus, also under more restrictive assumptions, the Boost for Mathematics appears to be a profitable investment.

F. Survey to teachers in Swedish

The survey to teachers in Swedish was directed to a representative sample of 5,000 teachers instructing in the Swedish language and/or social science in 500 primary and lower-secondary schools sampled between 2014/15–2017/18, with the purpose to assess the literacy instruction Swedish and social science. No other types of teachers or subjects were surveyed.

For our purpose we restrict attention teachers in the Swedish language, but it is important to note that about 65 percent of them also teach social science. Teachers instructing in both subjects are requested to provide an overall assessment of their literacy instruction in both these subjects.

We utilize the following two questions that specifically address different aspects of classroom activities for literacy instruction in the subjects of Swedish and/or social science. For all activities teachers are asked to report frequency as: very often, often, sometimes, rarely, never

1. *How often do you do the following in your teaching about and with texts?*

a) Choose texts based on students' interests
b) Select texts based on students' varying needs and prior knowledge
c) Clarify the purpose of your text choices
d) Allow students to read texts of their own choice
e) Reflect on critical aspects present in the tasks you present to students
f) Engage in discussions with students about the characteristics of language in different types of texts
g) Use questions to elicit reasoning and explanations from students
h) Adjust teaching based on students' responses and experiences
i) Model/demonstrate how students can approach a task
j) Have students work on conceptual understanding
k) Visit the school library/public library with your students
l) Provide individual feedback on students' work

m) Summarize what students should have learned
n) Teach the whole class simultaneously
o) Have students work together in small groups
p) Allow students to work individually

2 How often do you ask students to do the following in your teaching about and with texts?

a) Search for information in the text
b) Identify the main message of a text
c) Explain their own understanding of the text
d) Compare the text with their own experiences
e) Compare different texts
f) Predict what will happen in the text
g) Make generalizations and draw conclusions from a text
h) Describe the style and structure of the text
i) Pose their own questions to the text
j) Account for the author's perspective or opinions
k) Critically assess a text and its content
l) Discuss with each other what they have read
m) Discuss texts they have written themselves
n) Write to an authentic audience
o) Write a text together

In total, the two questions cover 31 different aspects of classroom activities. Answers for each activity is standardized in each wave of the survey, and is then aggregated as a) the mean value over all surveyed activities, and b) the first principal component over all surveyed activities.

References

Lindvall, Jannika, Ola Helenius, Kimmo Eriksson, and Andreas Ryve. 2022. "Impact and Design of a National-Scale Professional Development Program for Mathematics Teachers." *Scandinavian Journal of Educational Research* 66 (5): 744–59.

Gothenburg University. 1966. "Utvärdering genom uppföljning (Evaluation through follow-up), [database]", Department of Education.

Öckert, Björn. 2021. "School Absenteeism during the COVID-19 Pandemic – How Will Student Performance Be Affected?" In *Swedish Children and Youth during the COVID-19 Pandemic. Evidence from Research on Childhood Environment, Schooling, Educational Choice and Labour Market Entry*. IFAU Working Paper, 2021:3.

Skolinspektionen. 2012. "Lika för alla? Omrättning av nationella prov i grundskolan och gymnasieskolan under tre år."

———. 2021. "Ombedömning av nationella prov 2019." Diarienummer 2019:503.

Skolverket. 2016. "Slutredovisning av uppdrag att svara för utbildning." Mimeo Dnr 2011:643.

———. 2018. "The Boost for Mathematics. Slides from presentation in Karlstad."

———. 2015. "Trends in International Mathematics and Science Study (TIMSS) 2015, [database]."

Statistics Sweden. 2019a. "Flergenerationsregistret (The Multi-Generation Register), [database]."

———. 2019b. "Registerbaserad arbetsmarknadsstatistik (RAMS, Register-Based Labour Market Statistics), [database]."

———. 2019c. "Registret över befolkningens utbildning (UREG, The Education Register), [database]."

———. 2019d. "Registret över totalbefolkningen (RTB, The Population Register), [database]."

———. 1960-1990. "Folk- och bostadsräkning (FoB, Population and housing census", [database]."

———. 1968-1985. "Inkomst och taxering (IoT, Income and taxation register", [database]."