Scandinavian Journal of Educational Research

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/csje20

Strengthening Teachers in Disadvantaged Schools: Evidence from an Intervention in Sweden's Poorest City Districts

Caroline Hall, Martin Lundin & Kristina Sibbmark

To cite this article: Caroline Hall, Martin Lundin & Kristina Sibbmark (2022) Strengthening Teachers in Disadvantaged Schools: Evidence from an Intervention in Sweden's Poorest City Districts, Scandinavian Journal of Educational Research, 66:2, 208-224, DOI: 10.1080/00313831.2020.1788154

To link to this article: https://doi.org/10.1080/00313831.2020.1788154

Published online: 21 Jul 2020.

SCANDINAVIAN JOURNAL OF EDUCATIONAL RESEARCH
2022, VOL. 66, NO. 2, 208–224
https://doi.org/10.1080/00313831.2020.1788154

Routledge
Taylor & Francis Group

Strengthening Teachers in Disadvantaged Schools: Evidence from an Intervention in Sweden's Poorest City Districts

Caroline Hall[a,b], Martin Lundin [a,b] and Kristina Sibbmark[a]

[a]Institute for Evaluation of Labour Market and Education Policy (IFAU), Uppsala, Sweden; [b]Department of Economics, Uppsala Center for Labor Studies (UCLS), Uppsala, Sweden

ABSTRACT

Children growing up in disadvantaged neighborhoods perform much worse in school compared to children growing up under more favorable circumstances. We examine a program ("Coaching for Teaching") targeted at ten lower secondary schools in Sweden's most disadvantaged city districts. The program's aim was to enhance student performance by strengthening the teachers' professional development. Using a difference-in-differences design we estimate effects on several outcomes. We find a positive and statistically significant impact on student performance on standardized tests in English. Estimates for several other outcomes are also positive, but not statistically significant. For test scores in Swedish there is no improvement. An analysis of a survey among pupils supports the idea that the intervention enhanced the teaching and the classroom climate.

ARTICLE HISTORY
Received 18 December 2019
Accepted 27 April 2020

KEYWORDS
Disadvantaged schools; student performance; professional development; in-service teacher training; coaching; lower secondary schools; difference-in-difference

1. Introduction

Improving schooling outcomes among disadvantaged children and thereby reducing inequalities in life chances is a key challenge for public education systems worldwide. Children growing up in neighborhoods characterized by poverty, unemployment, poor health and crime tend to perform significantly worse in school than children growing up under more favorable circumstances (see e.g., Curto & Fryer, 2014; Sass et al., 2012). A major difficulty faced by schools in these neighborhoods is to attract and retain effective teachers (Chiang et al., 2017; Sass et al., 2012). Several studies find that teachers tend to prefer to work in schools that serve students from more well-off families, and that teacher turnover is higher in schools with more disadvantaged students (Hanushek et al., 2004; Karbownik, 2014; Lankford et al., 2002; Scafidi et al., 2007). At the same time, previous research identifies teacher quality as a crucial factor for enhancing student performance (e.g., Aaronson et al., 2007; Rivkin et al., 2005; Rockoff, 2004). A strategy to enhance teacher quality in disadvantaged schools could be to support the professional development among the existing pool of teachers. In-service training of teachers is a widespread practice around the world (e.g., Jacob & Lefgren, 2004a; Kraft et al., 2018), but evidence on the effectiveness of such programs in disadvantaged schools is scarce.

We study an intervention that aimed at enhancing student performance in disadvantaged neighborhoods through strengthening the teachers' professional development. The program, which started in the school year of 2012/13 and lasted three years, was targeted at ten poorly performing

CONTACT Martin Lundin martin.lundin@ifau.uu.se Institute for Evaluation of Labour Market and Education Policy (IFAU), P.O. Box 513, Uppsala SE-75120, Sweden; Department of Economics, Uppsala Center for Labor Studies (UCLS), P.O. Box 513, Uppsala SE-75120, Sweden

Supplemental data for this article can be accessed https://doi.org/10.1080/00313831.2020.1788154

lower secondary schools (grades 7–9) in Sweden's most disadvantaged city districts (often referred to as "urban development districts"). The targeted districts all have low employment rates, high reliance on social assistance, and a high share of pupils that do not manage to qualify for entering high school.

The intervention, which was called "Coaching for Teaching" (CFT) (*Handledning för lärande*), was initiated by the Swedish government as an effort to identify successful strategies for raising academic results in disadvantaged schools. The main component of the program was coaching, both in smaller groups and, if the teacher desired, individually. All teachers participated in group meetings, with a trained coach, every three weeks throughout the program period. The individual coaching was usually given at eight different occasions. A second important part of the program consisted of courses and seminars aiming to enhance the teachers' skills. For instance, almost all teachers participated in a quite extensive course on knowledge and language enhancing teaching strategies, which had the explicit purpose of improving the teaching of immigrant children with poor ability in the Swedish language.[1]

Out of the around 25 schools that were eligible for the intervention, based on being located in the targeted city districts, 10 were selected for participation. Although randomly assigning schools to treatment was not feasible, the selection process tried to ensure that treated and control schools would be balanced in terms of pre-treatment trends in student performance. We use a difference-in-differences design and rich register data to estimate the effect of the CFT intervention on students' results on standardized test in math, Swedish and English as well as on GPA and admission to upper secondary school.

Our results show a large and statistically significant positive impact on student performance on the standardized tests in English. The estimate for math is also positive and large, but not statistically significant at conventional levels; the same applies to GPA and admission to upper secondary school. There is no indication of improvement of test scores in Swedish. We also analyze a questionnaire with pupils, taken both before and towards the end of the intervention. These results indicate positive effects on students' perceptions of how their teachers teach and the classroom climate.

It is not possible to disentangle the precise mechanisms behind our findings. However, the positive impact on how students perceive the teaching indicate that the program indeed increased teacher skills and/or dedication. A qualitative analysis by Assadi et al. (2015) shows that the teachers' opinion about the usefulness of the group coaching varies a lot: some teachers are positive, but most are neutral or critical to this part of the program. One-to-one coaching is regarded as more valuable. Moreover, the teachers highly valued the course on knowledge and language enhancing teaching strategies. Hence, the results are in line with an interpretation that individual coaching and further training of teachers constitute valuable parts of the program.

Our paper is related to the broader literature on how to improve schooling outcomes in disadvantaged city districts (Cullen et al., 2013; Curto et al., 2011). While promising outcomes have been documented for several interventions – e.g., smaller classes (Chetty et al., 2011; Krueger, 1999; Krueger & Whitmore, 2001), smaller schools (Bloom et al., 2010), individualized math instruction (Cook et al., 2015), and mandatory summer school (Jacob & Lefgren, 2004b) – the success of many programs is likely to be contingent on the presence of effective teachers. In fact, the results in Angrist et al. (2013) indicate that selective hiring of teachers is one of the crucial components behind the success of several urban charter schools that serve many minority and low-income students in the US. A number of papers have examined the effectiveness of programs directed towards supplying disadvantaged schools with more effective teachers, e.g., through alternative qualification routes, with somewhat mixed findings on student performance.[2] Efforts to increase teacher effectiveness have also been conducted through offering teachers financial incentives based on how the pupils

[1]The intervention also contained additional components that in practice were less important (see Section 3).

[2]For instance, Chiang et al. (2017) find that math teachers from the alternative certification program "Teach for America" are more effective than other math teachers in high-poverty secondary schools in the US. Glazerman et al. (2006) find that also elementary school students with TFA teachers outperform students of other teachers in math, but not in reading, while Clark et al. (2015) find no significant difference on average in neither math nor reading.

perform. However, most evaluations in developed countries do not find evidence of enhanced student performance due to such incentives (Fryer, 2013; Goodman & Turner, 2010). Our study contributes to this discussion by providing empirical evidence on the effects of in-service training and professional development of teachers working in disadvantaged schools.

Our paper also adds to the literature on professional development of teachers. Quantitative evidence on the impact of in-service teacher training programs on student achievement is limited and the findings are mixed (Cilliers et al., 2019; Kraft et al., 2018; Popova et al., 2018). However, the literature focusing on one-to-one coaching of teachers is more solid; such initiatives often seem to generate rather positive (although often not that large) effects on student performance (Kraft et al., 2018).

2. Institutional Setting

2.1. The Swedish Education System

Sweden has nine years of compulsory schooling with a comprehensive curriculum. Children usually start school in the fall of the year they turn seven and finish 9th grade the year they turn 16. Traditionally, compulsory schooling has been organized in three stages: grades 1–3, 4–6 (lower vs. upper primary school) and 7–9 (lower secondary school). These days the organization is more flexible and the stages are no longer as distinct. After compulsory school, almost all pupils continue to upper secondary education; admission is based on students' 9th grade GPA.

2.2. Urban Development Districts

Like many other countries, Sweden has neighborhoods with high levels of social exclusion. To promote positive developments in such neighborhoods, the government directs specific support to city districts that are considered most disadvantaged, often referred to as "urban development districts". From 2012, 15 city districts with low employment rates, high reliance on social assistance and poor schooling outcomes were subject to such support.[3] During 2013−2014, the government devoted financial resources, amounting to SEK 100 million per year (approximately USD 11.6 million), to stimulate urban development in these areas.[4] They also launched the program "Coaching for Teaching", which we describe in detail in Section 3.

Poor schooling outcomes is one of the defining characteristics of the urban development districts. The first column of Table 1 displays average school results for pupils finishing compulsory school in these districts in 2012, while the second column displays the same averages for students graduating from other schools. We can see that students in the urban development areas have considerably lower GPA (165 vs. 210) as well as results on the standardized tests (percentile rank 38 vs. 50). They are also much less likely to have obtained the qualifications necessary to continue to upper secondary education. Their background characteristics are remarkably different: For instance, about 36 percent are born in another county, compared to 9 percent for the rest of the schools; and 78 percent have two parents born abroad, while the corresponding number for the other schools is 17 percent. The students are also much more likely to have recently immigrated to Sweden. Their parents have on average substantially lower earnings, education level, and are four times more likely to receive social assistance.

3. The "Coaching for Teaching" Program

The CFT intervention took place in ten public lower secondary schools during the three school years 2012/13–2014/15. The government provided financial resources and stated the program's overall

[3]The formal criteria for being considered an urban development district was having (i) an employment rate below 52 percent; (ii) a social assistance dependency rate above 4.8 percent; and (iii) less than 70 percent of students being qualified to enter upper secondary education. The districts also needed to have a population size exceeding 4,000. (Government decision A2012/174/IU).
[4]This implies around SEK 10,400 (USD 1200) per Swedish inhabitant.

Table 1. Characteristics of pupils graduating from compulsory school in urban development districts and all other districts, 2012.

	Schools in urban development districts	All other schools
GPA, 9th grade	164.7	210.4
Average results on the standardized test, 9th grade (percentile ranked)	37.7	50.2
Eligible for upper secondary education	0.61	0.88
Girls	0.52	0.49
Born in another country	0.36	0.09
Immigrated 0–4 years ago	0.14	0.03
Both parents are born in another country	0.78	0.17
Parents receive social assistance	0.35	0.08
Mother has post-secondary education	0.17	0.40
Father has post-secondary education	0.18	0.31
Mother's earnings (SEK)	119,119	252,575
Father's earnings (SEK)	164,429	358,898

Note: All variables originate from Statistics Sweden and are described in more detail in Section 4. Average test result refers to students' average grade on the standardized tests in math, English and Swedish.

orientation and aim, and the National Agency for Education (*Skolverket*) was given the task of formulating its more precise content.

The program's primary focus was to enhance teachers' professional development, with the goal of improving the instruction and thereby student performance. The intervention's main component was coaching. All teachers participated in coaching in small groups (3–12 participants per group) every three weeks throughout the program period. The idea was that teachers, through these group discussions, would strengthen their self-confidence and ability to handle the challenges of the teaching profession.[5] The coaches were especially qualified teachers hired and trained by the National Agency for Education. In addition, teachers that were interested could receive one-to-one coaching, usually at eight occasions.[6] The principals were also offered coaching in groups and individually, and most of them chose to participate (Assadi et al., 2015).

As a second crucial part of the program, the agency organized several courses and seminars with the aim of improving the teachers' skills. Nearly all teachers participated in an extensive course that had the aim of improving teaching towards pupils with poor ability in the Swedish language. This course focused on teaching strategies (including concrete pedagogical tools) to enhance knowledge and language development among immigrant pupils. The course was organized by the National Center for Swedish as a Second Language.[7] In addition, further training was offered on topics such as formative assessment, subjective didactics and motivational teaching methods (Assadi et al., 2015).

The schools were also given the opportunity to apply for extra funding for a variety of purposes: For organizing teaching outside regular school hours (e.g., homework assistance and teaching during school breaks), for training tutors that could assist newly arrived immigrant pupils in their native language (*studiehandledare*), and for efforts to improve communication between teachers and parents. Lastly, the National Agency for Education developed support material related to the various parts of the program; this material was made available to all participating schools (Assadi et al., 2015).

Table 2 shows how the resources were allocated: One third of the program budget was used for costs associated with the administration of the program, one third for coaching activities and the last

[5]The groups consisted of 3–12 teachers. The meetings usually lasted around 1.5 h. The teachers decided what to discuss (one question each meeting) and everything that was said in the discussions was confidential. A trained external coach chaired the meetings. Usually, the same group constellation was used throughout the program period (Assadi et al., 2015).

[6]Around 25 percent (94 individuals) of the teachers participated in individual coaching. The individual coaching usually took place for three months (Assadi et al., 2015).

[7]The National Center for Swedish as a Second Language is part of Stockholm University. The teachers participating in the course met eight work days spread out over a year. They also had assignments between the meetings. Note that this course was not targeted at improving Swedish. The aim was rather how improve the skills of how to teach children with poor abilities in Swedish in all subjects.

Table 2. Allocation of resources in the "Coaching for Teaching" intervention, 2012–2014.

	SEK	Percent
Group coaching	10,304,119	22.8
Ono-to-one coaching	2,457,724	5.4
Coaching of principals	1,787,500	4.0
Further training of teachers	8,157,595	18.0
Extra funding for a variety of purposes	7,466,701	16.5
Administration and implementation of the program	15,062,588	33.3
Total	45,245,227	100.0

Note: "Coaching of principals" also includes resources spent on further training of principals. Schools were able to apply for extra funding for the following purposes: organizing teaching outside regular school hours, for training certain tutors that could help newly arrived immigrants in their native language, and for improving communication between teachers and parents.

third was split almost evenly between further training and extra funding for a variety of purposes. Altogether, the resources used within the project correspond to around 1.4 extra teachers per school and year. The schools that participated in the program employed on average 35 teachers each, which means that an alternative use of the resources would have been to increase the teaching staff by approximately 5 percent.

A qualitative analysis of the CFT intervention, based on in-depth interviews with 34 teachers participating in the program, is included in Assadi et al. (2015). Also the principals at the participating schools were interviewed. The analysis shows that the program had a clear influence on the schools. Especially the various efforts to strengthening teachers' competence, through coaching and further training, affected daily work life throughout the program period. The opinions about the usefulness of the group coaching, which was the most visible part of the program, diverge. Some teachers found it to be a waste of time. Others, however, said that they had benefited from the group discussions, although most of them could not pinpoint more precisely what they had learned. The individual coaching and training activities – especially the course on knowledge and language enhancing teaching methods – were regarded as much more valuable.

As noted above, the schools could also apply for extra funding for activities such as homework assistance, teaching during school breaks and training of tutors. The results in Assadi et al. (2015) show that the schools did apply for extra funding, but that local practices did not change. For instance, the schools were already offering homework assistance and teaching during school breaks. We cannot conclude that these parts of the CFT program were unimportant, but it seems like activities aiming at improving teachers' professional development were the most prominent parts of the project.

3.1. The Selection of Schools for Participation

The National Agency for Education was given the task of selecting schools for participation. The program was directed towards public schools in urban development areas with poor academic results. We had the opportunity to advise the agency regarding the selection of schools, although they themselves made the final decision on the selection process.

In order to try to ensure there would exist a control group of schools with similar pre-treatment trends in student performance, we created ten pairs of schools (located in the same municipality) with similar developments in terms of GPA and eligibility for upper secondary education during the years preceding the intervention. Only one of the schools within each pair was then given the opportunity to participate. Although randomization within pairs would have been ideal for the purpose of evaluation, the National Agency for Education did not believe this to be a feasible strategy. The agency instead chose the school they found most suitable for participation within each pair. Reasons for why schools sometimes were considered less suitable included, e.g., lack of interest on the part of the principal, concerns that the school would be closed down in the near future, and awareness that the school recently had been the target of other types of

Table 3. Descriptive statistics for pupils who finished grade 9 at intervention, control and other schools 2009–2014.

	Intervention schools	Control schools[a]	Other schools
Age	16.1	16.0	16.0
Girl	0.487	0.485	0.487
Born in another country	0.372	0.343	0.087
Immigrated 0–2 years ago	0.092	0.092	0.014
Immigrated 3–4 years ago	0.064	0.057	0.013
Immigrated 5–6 years ago	0.054	0.048	0.013
Both parents are born in another country	0.758	0.783	0.160
Mother's level of education[b]			
Compulsory school or lower	0.363	0.380	0.102
Upper secondary education	0.333	0.350	0.472
Post-secondary education	0.186	0.149	0.392
Information is missing	0.118	0.121	0.034
Father's level of education[b]			
Compulsory school or lower	0.265	0.304	0.136
Upper secondary education	0.358	0.362	0.490
Post-secondary education	0.198	0.165	0.304
Information is missing	0.179	0.168	0.071
Mother's earnings (SEK)[b]	121,996	114,255	250,303
Father's earnings (SEK)[b]	166,728	154,243	355,175
Information about mother is missing	0.043	0.040	0.017
Information about father is missing	0.103	0.092	0.039
Number of individuals	4,810	4,274	623,697
Number of schools[c]	10	10	1,680

Note: All variables originate from Statistics Sweden, and are described in Section 4. [a]The control group consists of all public schools in the targeted city districts that did not participate in the intervention. [b]Parents' education and earnings are measured the year before the student finished 9th grade. Average earnings are computed among parents who were Swedish residents that year. [c]Number of "other schools" is measured in 2014.

interventions.[8] A couple of control schools in the initial pairs of schools in fact turned out to be located outside the boundaries of the targeted districts. These schools will be excluded from the analyses; we instead use all non-treated schools within the targeted districts as the control group.

The non-random selection of schools makes it possible that the two groups will differ in terms of average student characteristics (although it is not obvious whether we should expect the intervention schools to be positively or negatively selected). Table 3 compares the background characteristics of students enrolled in intervention schools (col. 1) and all remaining schools (col. 2) in the targeted districts. The table shows that the two groups are very similar on average, although the share of foreign-born students is slightly higher in the intervention schools as are the parents' education level and earnings. However, these differences are marginal in comparison to the striking differences relative to the rest of the schools in the country (col. 3).

4. Data and Research Design

Our analysis is based on individual level register data, which we have obtained from Statistics Sweden. To measure school performance, we use information on students' grades on the national standardized tests in Swedish, English and math, which are given during the spring semester of grade 9. All students take the same tests. The tests are corrected and graded locally at the schools, but there are detailed instructions from the National Agency for Education on how this should be done. To account for changes in grading standards over time, the students are percentile ranked within cohort based on their performance on each test. We also use information on the student's (percentile ranked) GPA, and whether he/she was admitted to a regular program in upper secondary

[8]One of the ten schools that initially were selected for participation closed shortly afterwards, and a new school was instead offered to participate. The new school was able to join the project already during the first semester and therefore received almost the same amount of treatment as the other intervention schools. The school that was shut down very early on is not included in the analysis.

school directly after finishing 9th grade as outcome variables. The GPA has the advantage of capturing effects for a broader range of subjects than those covered by the standardized tests. The obvious drawback is that it is a less objective measure of student performance.

We also have access to background variables: age, sex, each parent's education level and earnings, and immigrant background. To account for immigrant background in the analysis, we control for whether or not the pupil is born in another country, whether or not both of his/her parents are born in another country, and whether he/she recently immigrated to Sweden. The full list of background variables and descriptive statistics are displayed in Table 3.

To estimate the effect of the CFT intervention we compare how student performance develops for the schools that took part in the program, to the development for the untreated schools in the targeted districts, using a difference-in-differences design. Only public schools are included. The rationale for limiting the control group to untreated schools within the same city districts is twofold: (i) The credibility of the analysis hinges on the comparability of the two groups of schools. City districts that are considered urban development districts all have low employment rates, high reliance on social assistance, poor schooling outcomes, and the two groups of schools have very similar student characteristics (see Table 3). (ii) The urban development districts have also been the target of other initiatives that may affect educational outcomes (see Section 2.2). Thus, expanding the control group to include schools in other districts would make it harder to determine whether any observed differences in development were due to the CFT intervention or these other initiatives.

However, choosing a control group of schools from the same districts comes with a potential disadvantage: It is possible that the municipalities would choose to compensate the schools that were not selected for participation, e.g., with additional resources. Such a response could lead us to underestimate the impact of the CFT intervention on student performance.[9]

Our analysis is based on data 2009−2014. The start date is chosen to keep the panel of schools as balanced as possible (some of the schools are relatively new). The rationale for ending the evaluation period in 2014 – only two years after the start of the intervention – is that the government thereafter introduced an additional policy change that affected some of the schools in the sample.[10] Hence, after 2014 it is difficult to separate the effect of the CFT program from the effect of other policy changes. The obvious drawback with a short follow-up period is that the cohort that was most affected by the intervention (those finishing 9th grade in 2015), and potentially benefited the most, are not included in the analysis.

We estimate the following regression model:

$$y_{ist} = \alpha_0 + \beta_1(CFT_s{*}D_{2013}) + \beta_2(CFT_s{*}D_{2014}) + \sum_{t=2009}^{2014} \theta_t D_t + \boldsymbol{\delta X}_i + \gamma_s + e_{ist} \quad (1)$$

where i indexes individual, t the year the individual finished 9th grade, and s the school he/she attended. y_{ist} is the individual's (percentile ranked) grade on the standardized test in math, Swedish or English; GPA; or an indicator for admittance to upper secondary school. CFT_s is an indicator that takes the value one if the school the student attended participated in the intervention; otherwise it is zero. D_t are dummies for graduation year, $\boldsymbol{X}_i$ is a vector of individual and parental background characteristics (see Table 3), and γ_s represent school fixed effects. e_{ist} is the error term.

The parameters of interest, β_1 and β_2, are the difference-in-differences estimates of the effect of the intervention for the two cohorts of pupils that were treated – those who finished grade 9 in 2013 and those who finished in 2014. The first of these cohorts began 9th grade when the intervention was

[9]The same problem could occur if it is common that teachers who have received additional training through the intervention start working at untreated schools within the same city districts.

[10]In March 2014 the government decided to earmark central government grants for hiring additional teachers at some disadvantaged schools (Government Regulation 2014, p. 145). Pupils that graduated in the spring of 2014 (or earlier) should be unaffected by this policy change.

launched in the fall of 2012. However, the activities at the schools generally did not start until later during that semester, which means that they were just in place for about one semester before these students took the standardized tests (Assadi et al., 2015). Since the activities were directed towards the teachers, and not directly towards the students, this is probably a too short time period for there to be any noticeable effects on student performance. It is more likely that the intervention would affect the performance of the second cohort of students, those who finished in 2014. The activities at the schools were in place for about three semesters before this cohort of students took the standardized tests. Hence, if the CTF intervention was effective in raising student performance, we expect $\beta_2 > \beta_1$.

By incorporating school fixed effects, our model accounts for all unobserved differences across schools that remained constant over time. But to interpret β_1 and β_2 causally we need to assume that trends in student performance would not have differed between intervention and control schools in the absence of the CFT intervention. Figure 1, which plots average student outcomes for intervention and control schools (and also all other schools) during our study period, indicates that this is a realistic assumption in this case. For none of the outcomes we study are there any clear indications of divergent trends for treated and control schools during the period before the intervention started (2009−2012). In Section 5.2 we test this more formally.

Figure 1 suggests that performance in English may have improved among students graduating in 2014 (although the difference between the two groups of schools is not statistically significant in this simple year to year comparison). But we see no signs of improvements in math or Swedish. There is also a tendency for students to obtain higher GPAs following the intervention, but no clear indication of an increase in admission to upper secondary education.

5. Results

This section presents the results. We begin by showing results for the full sample of students (Section 5.1). This is followed by some robustness checks (Section 5.2), and thereafter we show results for different subgroups (Section 5.3). Last, we exploit a pupil questionnaire to shed some light on the mechanisms behind our findings (Section 5.4).

5.1. Main Results

Table 4 displays results for all five outcome variables. We show results both with and without individual background controls included in the model. To allow for correlation in error terms of individuals attending the same school, standard errors should be clustered at the school level. However, since we rely on rather few clusters (20 schools), cluster-robust standard errors may be underestimated. To make correct inference with few clusters, we instead use the wild bootstrap method as suggested by Cameron et al. (2008) and Cameron and Miller (2015). Table 4 (and all following tables) displays the *p*-values resulting from the wild bootstrap procedure.

As expected, given the short time of exposure, there is no indication that the CFT intervention affected students finishing 9th grade in 2013. Among those who graduated in 2014, we see improved results on the standardized test in English among students attending intervention schools. This effect is statistically significant at the 5 percent level, and the estimate is robust to the inclusion of background controls in the model. In terms of magnitude, a 6.4 percentile points increase (col. 2) is substantial. It implies that the gap in test results between students in urban development districts and other schools in the country (see Figure 1) is closed by approximately 50 percent. The point estimates are positive and large also for some of the other outcomes variables, but they are all imprecisely estimated and for the most part not significantly different from zero at conventional levels of significance.

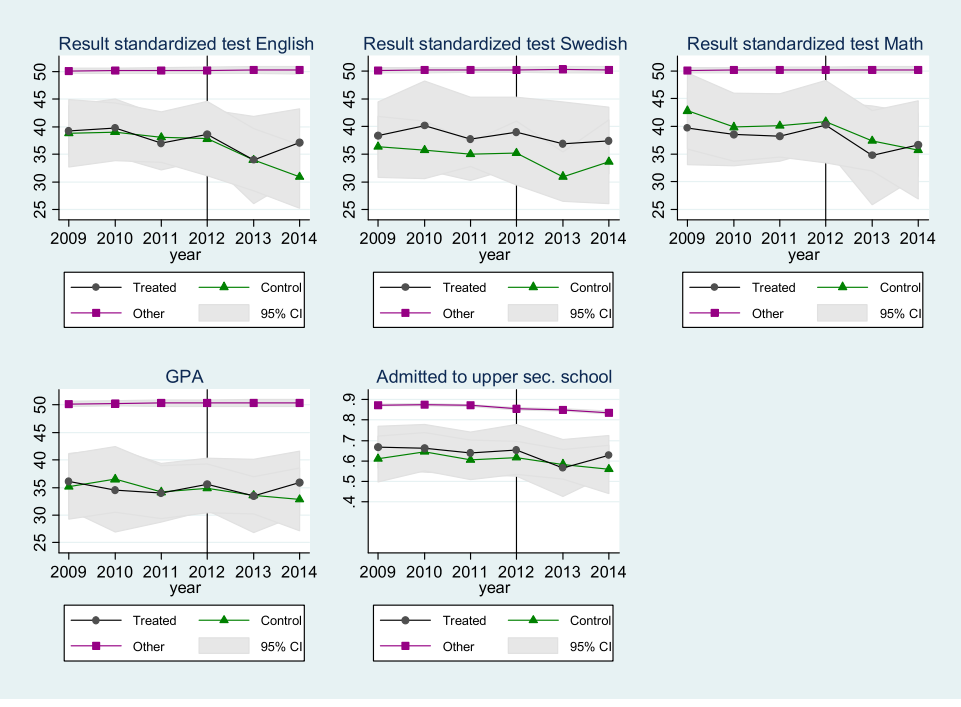

Figure 1. Results on standardized tests, GPA, and admission to upper secondary school for intervention schools, control schools and all other schools, 2009–2014. Note: Results (grades) on the standardized tests and GPA are percentile ranked within cohort. The standard errors allow for clustering by school. The vertical line marks the year before the CFT intervention started.

5.2. Robustness Checks

A causal interpretation of our estimates relies on the assumption that trends in student performance would not have differed between intervention and control schools in the absence of the intervention. The patterns in Figure 1 gave us no reason to doubt this assumption. To more formally assess this, we estimate a number of placebo-regressions: We estimate the same model, but artificially set the intervention year to one, two and three years before the actual start year (see Table 5). Any significant placebo-estimates for these years would cast doubts on the validity of the identifying assumption. Reassuringly, we find no statistically significant placebo-estimates.

The fact that our sample consists of only 20 schools means that the development of an individual school potentially can have a large influence on the results. This raises concerns that the estimated effect on English test scores would be largely driven by improvements in a particular school. To investigate the sensitivity of our results, we re-estimate all regressions excluding one school at a time. The estimate for English is stable across these 20 regressions; see Table A1 in the online appendix. The patterns for the other outcomes are, for the most part, also unaffected.

5.3. Heterogenous Effects?

Some of the intervention's activities explicitly focused on enhancing the teaching of immigrant children with poor ability in the Swedish language. It is therefore of special interest to examine how the intervention impacted students that recently immigrated to Sweden. Table 6 shows results from separate regressions for students that immigrated 0–4 years ago and all other students.

The estimated effect on test results in English, for the 2014 graduation cohort, is similar in size for newly arrived immigrant children and others (although it is less precisely estimated for the former group); for all other outcomes the estimates are smaller in size for the newly arrived and sometimes

Table 4. Effects of the CFT intervention on results on standardized tests, GPA, and admission to upper secondary school (*p*-values in parentheses).

A. Test results	English		Swedish		Math	
	(1)	(2)	(3)	(4)	(5)	(6)
Effect 2013	0.453	1.512	3.288	3.548	1.054	1.597
	(0.873)	(0.590)	(0.280)	(0.237)	(0.818)	(0.712)
Effect 2014	6.640*	6.393**	2.218	1.170	5.615	5.538
	(0.052)	(0.040)	(0.610)	(0.784)	(0.326)	(0.326)
Controls	No	yes	no	yes	no	yes
Observations	7,140	7,139	7,296	7,295	7,087	7,086
R-squared	0.073	0.214	0.072	0.226	0.078	0.169
Mean of outcome variable	37.182	37.184	36.522	36.526	38.785	38.790
B. GPA and admission to upper secondary school	GPA		Admitted			
	(7)	(8)	(9)	(10)		
Effect 2013	1.084	2.011	−0.036	−0.017		
	(0.672)	(0.428)	(0.453)	(0.721)		
Effect 2014	5.109*	4.246	0.067	0.059		
	(0.058)	(0.118)	(0.312)	(0.350)		
Controls	no	yes	no	yes		
Observations	9,084	9,082	9,084	9,082		
R-squared	0.056	0.273	0.083	0.290		
Mean of outcome variable	34.753	34.759	0.621	0.621		

Notes: Results (grades) on standardized tests and GPA are percentile ranked within cohort. All regressions control for year and school fixed effects. Col. (2), (4), (6), (8) and (10) additionally control for sex, age, born in another country, years since immigration (4 categories), parents born in another country, mother's and father's education level (4 categories), and mother's and father's (log) earnings (see Table 3 for details). *P*-values (in parentheses) are obtained using the wild bootstrap method. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

negative. Thus, we find nothing to suggest that the CFT intervention primarily benefitted pupils with poor ability in the Swedish language; if anything, the pattern is the reverse. Among pupils that have resided in Sweden longer than four years, we see a positive and statistically significant effect of the intervention on the GPA, indicating that the intervention benefitted students' performance in more subjects than those covered by the standardized tests.

To further investigate whether the intervention seems to primarily have benefitted relatively weak or strong pupils, we estimate effects on the probability of achieving a GPA above the

Table 5. Placebo-tests (*p*-values in parentheses).

	(1) English	(2) Swedish	(3) Math	(3) GPA	(4) Admitted
Placebo estimate	1.790	0.687	2.746	2.004	0.029
2012	(0.455)	(0.860)	(0.298)	(0.181)	(0.364)
Observations	4,874	5,000	4,863	6,224	6,224
R-squared	0.195	0.218	0.167	0.282	(0.308)
Placebo estimate	−1.233	−1.150	0.736	0.863	0.001
2011	(0.467)	(0.798)	(0.745)	(0.670)	(0.984)
Observations	3,657	3,738	3,671	4,732	4,372
R-squared	0.198	0.217	0.175	0.293	0.318
Placebo estimate	1.143	4.372	2.466	−0.884	−0.006
2010	(0.480)	(0.233)	(0.496)	(0.685)	(0.888)
Observations	2,539	2,629	2,591	3,240	3,240
R-squared	0.206	0.239	0.188	0.309	0.336

Notes: Results (grades) on standardized tests and GPA are percentile ranked within cohort. All regressions control for year, school fixed effects and the same control variables as in Table 4. *P*-values are obtained using the wild bootstrap method. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 6. Effects of the CFT intervention for new immigrants and others (*p*-values in parentheses).

A. Test results	English		Swedish		Math	
	New immigrant	Others	New immigrant	Others	New immigrant	Others
Effect 2013	1.741	1.127	−0.567	3.446	−0.567	1.747
	(0.761)	(0.708)	(0.960)	(0.231)	(0.960)	(0.697)
Effect 2014	6.670	6.032*	−2.954	1.338	−2.954	6.808
	(0.265)	(0.078)	(0.680)	(0.792)	(0.680)	(0.268)
Controls	yes	yes	yes	yes	yes	yes
Observations	522	6,617	684	6,746	684	6,402
R-squared	0.209	0.189	0.193	0.204	0.193	0.172
Mean of outcome var.	18.501	38.658	18.168	38.020	31.148	39.606
B. GPA and admission upper sec. school	GPA		Admitted			
	New immigrant	Others	New immigrant	Others		
Effect 2013	−1.122	2.173	−0.009	−0.025		
	(0.758)	(0.418)	(0.881)	(0.649)		
Effect 2014	−4.725	6.162**	0.015	0.068		
	(0.302)	(0.039)	(0.852)	(0.308)		
Controls	yes	yes	yes	yes		
Observations	1,388	7,694	1,388	7,694		
R-squared	0.231	0.199	0.216	0.163		
Mean of outcome var.	12.610	38.755	0.170	0.702		

Notes: Results (grades) on standardized tests and GPA are percentile ranked within cohort. "Newly arrived" is defined as children that immigrated 0–4 years ago. All regressions control for year, school fixed effects and the same control variables as in Table 4. *P*-values are obtained using the wild bootstrap method. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 7. Effects on the probability of receiving a GPA above the 25th, 50th, and 75th percentile in the GPA distribution (*p*-values in parentheses).

	GPA> 25th percentile	GPA> median	GPA> 75th percentile
Effect 2013	0.011	0.048	0.037
	(0.814)	(0.206)	(0.178)
Effect 2014	0.022	0.079*	0.059*
	(0.467)	(0.076)	(0.062)
Controls	yes	yes	yes
Observations	9,082	9,082	9,082
R-squared	0.293	0.216	0.149
Mean of outcome variable	0.752	0.500	0.248

Notes: GPA is percentile ranked within cohort. All regressions control for year, school fixed effects and the same control variables as in Table 4. *P*-values are obtained using the wild bootstrap method. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

25th, 50th, and 75th percentile of the grade distribution (see Table 7). The results are in line with those in Table 6 as they suggest that the intervention primarily benefited students in the middle and upper part of the grade distribution. We find a positive and statistically significant effect on the probability of receiving a GPA above the median as well as above the 75th percentile, while the estimate for the 25th percentile is much smaller and statistically insignificant. However, it is important to point out that pupils in the middle-upper part of the grade distribution for these schools are still relatively weak in relation to the grade distribution for all students in the country.

Last, we have analyzed whether the impact of the intervention differs for boys and girls; see Table A2 in the online appendix. We find no indications of gender differences in effects.

5.4. Effects on School Practices

To learn more about the effects of the CFT program we make use of a survey of pupils. The control group (which is not identical to the control group used in the previous analyses) was constructed to match the intervention schools in terms of location (municipality) and pre-intervention trends in

GPA and eligibility for upper secondary education. The treatment and control group here consist of the initial pairs of schools described in Section 3.1. The survey was taken before the intervention started and then at the end of the program. Although all ten pairs of schools answered the first round of the survey, only seven pairs can be included in the analysis: One of the schools that initially was selected for participation closed down just before the intervention started, but after the first round of the survey. The National Agency for Education then decided to include one of the comparison schools in the project. Furthermore, one of the remaining comparison schools refused to participate in the second round of the survey.[11] The two groups of schools are comparable in terms of average student characteristics in the beginning of the intervention (see Table A3 in the online appendix).

We approached all pupils in grade 7–9, and we received a response rate of 75 percent in the first survey round and 82 percent in the second. The average response rates were similar for intervention and comparison schools. To examine if the program affected students' responses, we again use a difference-in-differences strategy.

The main idea of the program was to strengthen the teachers' professional skills and confidence through coaching, collegial learning and further training. As a consequence, the instruction as well as the classroom climate was expected to improve, which should enhance learning and performance among the students. In line with this chain of events, we anticipate positive effects of the intervention on students' opinions about their teachers' ability to teach and the learning environment in the classroom. The survey therefore included several questions on students' perception of how their teachers teach and the classroom climate.

We use factor analysis of six survey items to empirically examine whether it is possible to identify a common underlying dimension of opinions about *teacher ability* in the responses (see Table A4 in the online appendix). For instance, we asked the students to what extent they think their teachers are good at teaching and creating an interest among the students. All six items load positively and high on a single dimension. These findings indicate that an index based on the items is likely to be a valid measure of students' perceptions of teacher ability. We use the factor scores from the analysis to compute an index standardized to have a mean of 0 and a standard deviation of 1. A higher score implies a more positive assessment of teacher ability. We also use the overall question: "How many of your teachers are good at teaching?" as an alternative outcome.

We employ a similar procedure to construct an index of opinions about the *learning environment* in the classroom. Three survey items capture the extent to which pupils find that the classroom climate is favorable for learning. The items load positively and high on a single dimension (see Table A5) and we construct an index (mean 0, standard deviation 1) where a higher score implies a more positive view of the classroom climate.

Table 8 shows the results when we use students' opinions of teacher ability as the dependent variable. In columns 1–3, the full index of six items is the outcome, whereas the single survey question is the outcome in columns 4–6. We estimate models without control variables (col. 1 and 4) and models controlling for background characteristics (col. 2 and 5). In the most restrictive specification, we also include school fixed effects (col. 3 and 6).

We find positive estimates in all specifications, indicating that the intervention improved students' perceptions about their teachers' ability to teach. The estimates suggest an increase by around 0.15 of a standard deviation when the index is used as the outcome, and an increase of about 0.2 on a scale 1–5 when the single survey item is used as outcome. The estimate becomes marginally

[11]The seven treatment schools included in the analysis of the questionnaire are a subset of the ten schools used as treatment schools when we analyze register data. Table A7 in the online appendix displays descriptive statistics of the samples used in the main analysis vs. the survey analysis. The students in intervention schools and control schools are quite similar. However, the control schools included in the survey is to some extent less negatively selected than the control schools included in the main analysis. Note that Table A7 and Table A3 should not be compared directly, since the samples of pupils are different (grade 9 vs. grade 7–9), the data source is not the same (questionnaire vs. register data) and the variables are not measured in the same way.

Table 8. Effects of the CFT intervention on students' perceptions of teacher ability (*p*-values within parentheses).

	Index			One item		
	(1)	(2)	(3)	(4)	(5)	(6)
Effect of CFT	0.199**	0.170**	0.139	0.172***	0.210***	0.174***
	(0.036)	(0.026)	(0.109)	(0.010)	(0.000)	(0.002)
Controls	no	yes	yes	no	yes	yes
School fixed effects	no	no	yes	no	no	yes
Observations	4,204	3,599	3,599	5,069	4,256	4,256
R-squared, adj.	0.005	0.098	0.124	0.003	0.054	0.087
Mean of outcome var.	0.000	0.000	0.000	3.769	3.769	3.769

Note: In models 1–3, an index of perceived teacher ability based on six survey items from the student survey is used as the dependent variable. The index is computed using the factor loadings from a factor analysis of the questions. The index is set to have a mean of 0 and a standard deviation of 1, and a higher score implies a more positive assessment of teacher ability (see Table A3 for details). In models 4–6, the following survey question is used as the dependent variable: "How many of your teachers are good at teaching?" The response scale is 1–5, where 1 = "none of them" and 5 = "all of them". Models (2), (3), (5) and (6) include the following control variables: sex, age, Swedish as native language (yes/no), have attended Swedish school for less than four years (yes/no), attends a class for newly arrived immigrants (yes/no), father's educational level (5 levels), mother's educational level (5 levels). *P*-values are obtained using the wild bootstrap method. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 9. Effects of the CFT intervention on perceptions of the learning environment in the classroom (*p*-values within parentheses).

	(1)	(2)	(3)
Effect of CFT	0.078	0.133***	0.093*
	(0.106)	(0.006)	(0.056)
Controls	no	yes	yes
School fixed effects	no	no	yes
Observations	4,739	4,066	4,066
R-squared, adj.	0.002	0.011	0.058
Mean of outcome variable	0.000	0.000	0.000

Notes: The dependent variable is an index computed from the factor loadings from a factor analysis of three questions in the student survey. This index is set to have a mean of 0 and a standard deviation of 1, and a higher score implies a more positive assessment of the learning environment in the classroom (see Table A4 for details). Models (2) and (3) include the following the same control variables as in Table 8. *P*-values are obtained using the wild bootstrap method. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

insignificant (p-value = 0.109) in the most restrictive specification when we use the index as the outcome (col. 3). However, we can only include pupils who have answered all six survey questions, and where we have information on all control variables, in this analysis. This decreases the sample size and therefore statistical precision.[12]

Table 9 displays the results when we study student perceptions of the classroom climate. The positive estimates suggest that the situation in the classroom became more favorable for learning due to the intervention. The estimate is marginally insignificant (p-value = .106) in the model without control variables, but statistically significant when covariates and/or school fixed effects are added to the model. In sum, the analysis suggests that the CFT intervention may have improved the teaching as well as the classroom climate at the schools, in line with the stated intentions.

As outlined in Section 3, the CFT program included some additional components. Most notably, schools could apply for extra funding to enhance teaching outside of regular school hours (e.g., homework assistance) and for tutors that could assist immigrant students in their native language. Table 10 shows estimates from regressions where we use responses to six survey questions related to these aspects as outcomes.

We find no indication that the intervention affected homework assistance and other forms of teaching outside of regular school hours (col. 1−4). But we find a rather large positive estimate when we study the presence of tutors that could assist students in their native language. The

[12]As a robustness check, we have estimated models with each survey question as the outcome separately (see Table A6 in the online appendix). The results are robust.

Table 10. Effects of the CFT intervention on students' perceptions of homework assistance, teaching outside regular school hours and help with school work in native language (*p*-values within parentheses).

	Receives homework assistance? (1)	Minutes per week of homework assistance? (2)	Pleased with homework assistance? (3)	Receives teaching outside regular school hours? (4)	Receives help with school work in native language? (5)	Pleased with help with school work in native language? (6)
Effect of CFT	−0.102	−15.944	0.005	0.015	0.071**	0.138
	(0.170)	(0.266)	(0.934)	(0.780)	(0.040)	(0.232)
Controls	yes	yes	yes	yes	yes	yes
School FE	yes	yes	yes	yes	yes	yes
Observations	4,205	3,513	1,692	4,132	3,057	613
R-squared, adj.	0.126	0.215	0.037	0.048	0.080	0.019
Mean of outcome var.	0.417	47.819	2.939	0.135	0.236	3.312

Notes: The outcomes in columns (1), (4) and (5) are dichotomous (0/1 = no/yes). The outcomes in columns (3) and (6) are measured on an attitude scale 1–4, where 4 = "very pleased". All models include school fixed effects and the same control variables as in Table 8. Findings are robust if these variables are excluded from the models, with the exception that the estimate in column (1) becomes slightly larger (−0.140) and statistically significant (*p*-value = 0.048) if no controls are added. *P*-values are obtained using the wild bootstrap method. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

probability that students born abroad answered that they receive help in their native language increased by 7 percentage points (col. 5). This seems like a quite sizeable effect (corresponding to around 30 percent of the mean). However, the fact that positive impacts on student performance are primarily found among students who have resided in Sweden longer than four years (see Table 6), suggests that improved access to this type of assistance is unlikely to account for the overall improvement in school results at the intervention schools.

As discussed in the introduction, a major challenge for schools located in disadvantaged city districts lies in attracting and retaining teachers. A possible explanation for the positive impact on student performance is that the schools, through the program, became more attractive workplaces and thereby, to a higher extent, managed to keep their teaching staff. We have access to data on teachers, and where they work, through the Teacher register (*Lärarregistret*). Using these data, we estimate the same type of difference-in-differences model as in the main analysis, but where the outcome is the probability of remaining employed at the same school the following school year. We find no effect on teacher mobility of the CFT intervention (see Table A8 in the online appendix).

6. Conclusion

Children growing up in neighborhoods with high levels of social exclusion in the form of unemployment, poverty and reliance on social assistance tend to perform significantly worse in school compared to children growing up under more favorable conditions. In this paper we have examined a public program, "Coaching for Teaching", initiated by the Swedish government to improve school results in ten disadvantaged lower secondary schools. The program consisted of several components, but primarily focused on teachers' professional development.

Our study contributes to research on how to improve schooling outcomes in disadvantaged city districts, and among pupils with disadvantaged background more generally (e.g., Angrist et al., 2013; Chiang et al., 2017; Cullen et al., 2013; Curto et al., 2011). Previous research has shown that teacher quality is a crucial factor for enhancing student performance (e.g., Aaronson et al., 2007; Rivkin et al., 2005; Rockoff, 2004), but also that it is difficult for schools in disadvantaged neighborhoods to attract and retain effective teachers (Chiang et al., 2017; Sass et al., 2012). We have added to the existent literature by presenting evidence suggesting that programs aiming to strengthen teachers' professional development might be a promising way forward for these schools.

We find a sizeable and statistically significant positive effect of the program on student performance on standardized tests in English. The estimates for students' test scores in math, GPA, and admission to upper secondary school are also positive and large, but not statistically significant at conventional levels. The low number of schools participating in the intervention may explain why the effects are not precisely estimated. For test scores in Swedish there is no indication of improvement.

Given that a course on teaching strategies to enhance knowledge and language development among immigrant pupils constituted an important part of the training package for teachers, it may seem remarkable that the positive effects on student performance are seen in other subjects than Swedish. However, students who do not have Swedish as their native language generally take the course "Swedish as a second language" rather than regular Swedish.[13] Since teachers of this subject already have training particularly aimed at teaching this group of students, it is perhaps not surprising that the effects are larger in other subjects.

We have also investigated whether the program primarily benefitted relatively weak or strong pupils. We find some indication of stronger effects for children who have lived in Sweden for more than four years. For this group we also see a positive and statistically significant impact on the GPA. The effects are also more accentuated for students in the middle and upper part of the grade distribution in the sample. However, it is essential to emphasize that these pupils are still relatively weak in relation to the grade distribution for all students in the country. Thus, it seems like weak pupils, but not necessarily the weakest, also benefitted from the program.

Furthermore, we have analyzed data from a student questionnaire. The results indicate that students' perceptions of the learning environment in classroom and their teachers' ability to teach improved. Hence, the program may, in line with its intentions, have enhanced the learning environment at the schools. Taken together, we interpret the empirical results as quite promising in the short run.

Like many other programs designed to fight social exclusion, the CFT program consisted of several components. This implies that it is hard to determine which parts of the program that were more and less successful. Coaching of teachers in small groups constituted the main part of the program. However, qualitative results in Assadi et al. (2015) warn against adopting the specific coaching model used within the project. According to that study, one-to-one coaching and further training were more promising parts of the program. Note that one-to-one coaching of teachers also seems to have generated quite promising results in other settings (Kraft et al., 2018). We think that it is appropriate to interpret our results as indications that strategies to strengthen teachers' professional development – without pinpointing the exact design of such a policy – can be a fruitful way to enhance student performance in disadvantaged areas.

We have only been able to examine short run effects. However, an important idea behind the program was to generate a more lasting impact compared to temporary increases in resources. For instance, while the impact of additional teachers is likely to disappear when resources are removed, increased professional competence among existing teachers might produce positive effects for subsequent cohorts even when the program is over. This idea builds on at least two important assumptions: First, that the teachers will remain employed at disadvantaged schools, or that improved teaching strategies can be transferred to new recruits. Second, the effects cannot simply be driven by enthusiasm over the fact that the schools were selected to participate in a project. We leave to future research to explore these issues.

Finally, we would like to emphasize some lessons to be learned for policy makers who are interested in initiating similar efforts at schools and who would like these efforts to be evaluated. The fact that we find large but imprecisely estimated effects for several outcomes shows that there is a need for further examinations of the potential of this type of initiatives. But it also illustrates the difficulty in determining effects when relatively few schools have the possibility to participate. Few participating schools risk leading to imprecise estimates, making it hard to determine if the program was effective

[13]All students still take the same national standardized test in Swedish.

or not unless the effects are of substantial size. Expanding the program to more schools would help to overcome this problem.

An additional important aspect to consider for future similar projects is the trade-off between the number of components included and the possibility of determining which components that were more and less successful. A project that affects many parts of the school has, of course, a greater potential of making a real impact, but this will come at the cost of it being difficult to pinpoint which parts that worked well and which parts that did not. One way forward is to, as we do here, rely on qualitative studies to shed some light on potential mechanisms. Another way would be to consider reducing the number of components involved in the same intervention.

Acknowledgements

We are grateful to Anna Sjögren, Olof Åslund, and Björn Öckert as well as seminar participants at IFAU and the IFAU-UCLS conference "How to improve school results? Policy design, implementation and evaluation" for valuable comments.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

This research did not receive any specific grant for funding agencies in the public, commercial, or not-for-profit sectors.

ORCID

Martin Lundin http://orcid.org/0000-0001-7379-9704

References

Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics*, *25*(1), 95–135. https://doi.org/10.1086/508733

Angrist, J. D., Pathak, P. A., & Walters, C. R. (2013). Explaining charter school effectiveness. *American Economic Journal: Applied Economics*, *5*(4), 1–27. https://doi.org/10.1257/app.5.4.1

Assadi, A., Hall, C., Lundin, M., & Sibbmark, K. (2015). *Erfarenheter och effekter av satsningar på lärare i skolor med låga elevresultat* (Report 2015:23). IFAU.

Bloom, H. S., Thompson, S. L., Unterman, R., Herlihy, C., & Payne, C. F. (2010). *Transforming the high school experience: How New York City's new small schools are boosting student achievement and graduation rates* (Report). Manpower Demonstration Research Corporation.

Cameron, C. A., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics*, *90*(3), 414–427. https://doi.org/10.1162/rest.90.3.414

Cameron, C. A., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, *50*(2), 317–372. https://doi.org/10.3368/jhr.50.2.317

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from project STAR. *Quarterly Journal of Economics*, *126*(4), 1593–1660. https://doi.org/10.1093/qje/qjr041

Chiang, H. S., Clark, M. A., & McConnell, S. (2017). Supplying disadvantaged schools with effective teachers: Experimental evidence on secondary math teachers from teach for America. *Journal of Policy Analysis and Management*, *36*(1), 97–125. https://doi.org/10.1002/pam.21958

Cilliers, J., Fleisch, B., Prinsloo, C., & Taylor, S. (2019). How to improve teaching practice? An experimental comparison of centralized training and in-classroom coaching. *Journal of Human Resources*. https://doi.org/10.3368/jhr.55.3.0618-9538R1.

Clark, M. A., Isenberg, E., Liu, A. Y., Makowsky, L., & Zukiewicz, M. (2015). *Impacts of the Teach for America investing in innovation scale-up*. Princeton, NJ: Mathematica Policy Research.

Cook, P., Dodge, K., Farkas, G., Roland, G., Fryer, J., Guryan, J., Ludwig, J., Mayer, S., Pollack, H., & Steinberg, L. (2015). *Not too late: Improving academic outcomes for disadvantaged youth*. Institute for Policy Research, Northwestern University.

Cullen, J. B., Levitt, S. D., Robertson, E., & Sadoff, S. (2013). What can be done to improve struggling high schools? *Journal of Economic Perspectives*, *27*(2), 133–152. https://doi.org/10.1257/jep.27.2.133

Curto, V. E., & Fryer Jr., R. G. (2014). The potential of urban boarding schools for the poor: Evidence from SEED. *Journal of Labor Economics*, *32*(1), 65–93. https://doi.org/10.1086/671798

Curto, V. E., Fryer Jr., R. G., & Howard, M. L. (2011). It may not take a village: Increasing achievement among the poor. In G. J. Duncan, & R. J. Murnane (Eds.), *Whither opportunity? Rising inequality, schools, and children's life chances* (pp. 483–506). Russel Sage Foundation.

Fryer, R. G. (2013). Teacher incentives and student achievement: Evidence from New York City public schools. *Journal of Labor Economics*, *31*(2), 373–407. https://doi.org/10.1086/667757

Glazerman, S., Mayer, D., & Decker, P. (2006). Alternative routes to teaching: The impacts of teach for America on student achievement and other outcomes. *Journal of Policy Analysis and Management*, *25*(1), 75–96. https://doi.org/10.1002/pam.20157

Goodman, S., & Turner, L. (2010). *Teacher incentive pay and educational outcomes: Evidence from the NYC bonus program* (Working Paper No. 10–07). Program on Education Policy and Governance.

Government decision A2012/174/IU, *Kommuner som ska omfattas av urbant utvecklingsarbete från 2012*.

Government Regulation. (2014:145). Förordning om extra statsbidrag till skolhuvudmän som inrättar karriärsteg för lärare i förskoleklasser och grundskolor i utanförskapsområden.

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2004). Why public schools lose teachers. *Journal of Human Resources*, *39* (2), 326–354. https://doi.org/10.2307/3559017

Jacob, B., & Lefgren, L. (2004a). The impact of teacher training on student achievement: Quasi-experimental evidence from school reform efforts in Chicago. *Journal of Human Resources*, *39*(1), 50–79. https://doi.org/10.2307/3559005

Jacob, B., & Lefgren, L. (2004b). Remedial education and student achievement: A regression-discontinuity analysis. *Review of Economics and Statistics*, *86*(1), 226–244. https://doi.org/10.1162/003465304323023778

Karbownik, K. (2014). *Do changes in student quality affect teacher mobility? Evidence from an admission reform* (Working Paper 2014:15). Institute for evaluation of labour market and education policy (IFAU). Uppsala.

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research*, *88*(4), 547–588. https://doi.org/10.3102/0034654318759268

Krueger, A. B. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics*, *114* (2), 497–532. https://doi.org/10.1162/003355399556052

Krueger, A. B., & Whitmore, D. M. (2001). The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from project star. *Economic Journal*, *111*(468), 1–28. https://doi.org/10.1111/1468-0297.00586

Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis*, *24*(1), 37–62. https://doi.org/10.3102/01623737024001037

Popova, A., Evans, D. K., & Arancibia, V. (2018). *Training teachers on the job. What works and how to measure it* (Policy Research Working Paper 7834). World Bank.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*(2), 417–458. https://doi.org/10.1111/j.1468-0262.2005.00584.x

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, *94*(2), 247–252. https://doi.org/10.1257/0002828041302244

Sass, T. R., Hannaway, J., Xu, Z., Figlio, D. N., & Feng, L. (2012). Value added of teachers in high-poverty schools and lower poverty schools. *Journal of Urban Economics*, *72*(2-3), 104–122. https://doi.org/10.1016/j.jue.2012.04.004

Scafidi, B., Sjoquist, D. L., & Stinebrickner, T. R. (2007). Race, poverty, and teacher mobility. *Economics of Education Review*, *26*(2), 145–159. doi:10.1016/j.econedurev.2005.08.006