EXPLORING THE ANCILLARY BENEFITS OF RESIDENCY AND DIFFERENTIAL STAFFING PROGRAMS:

REPORT PREPARED FOR THE OVERDECK FAMILY FOUNDATION

PRATIKSHYA BOHRA-MISHRA

REBECCA CASCIANO

JENNIFER PUMA

GLASS FROG SOLUTIONS

ACKNOWLEDGEMENTS

The authors thank the Overdeck Family Foundation for providing the vision, resources, and guidance to execute this study. We would also like to thank each of the residency program partners and their respective school partner organizations, as well as the differential staffing program partners for their participation in this study. This study would not have been possible without the support of the organizations and experts that helped us learn more about their programs, and collect the necessary data for the analyses. The National Center for Teacher Residency was particularly supportive and collaborative throughout the project. Others such as Public Impact, David Rosenberg at Education Resource Strategies, and Cynthia Osborne at University of Texas – Austin also offered valuable background and insights. Senad Lekpek provided considerable research assistance on this project. John Friedman provided valuable technical consultation. Please direct questions to Rebecca Casciano at rebecca@glassfrog.us.

CONTENTS

EXECUTIVE SUMMARY

This report was prepared for the Overdeck Family Foundation (Overdeck). It was guided by a set of research questions that will inform our understanding of the ancillary benefits of teacher residency and differential staffing (DS) programs. The report primarily examines two largely unexplored research questions:

- What is the added value of hosting residency personnel in a classroom, as measured by improved teacher effectiveness scores (TESes) during the year of the residency?
- Are teachers who worked with DS programs prior to becoming teachers more effective (as measured by student academic outcomes) than teachers who took more traditional routes to becoming teachers?

The report focuses on programs' impacts on outcomes that are secondary to their primary outcome goals, which, in the case of residency programs, includes training future teachers and, in the case of DS programs, includes improving student outcomes. The findings presented herein do not reflect the programs' effectiveness in terms of meeting their primary outcome goals. Thus, consumers of this report should not use the report to make judgements about the effectiveness of these programs at achieving their primary goals.

Residency programs are teacher preparation programs that provide experiential training to residents by placing them in classrooms with host teachers, as a part of a comprehensive graduate-level curriculum. We explore whether, by placing an additional adult (i.e., a resident) in the classroom, there are also positive, short-run benefits for student learning. We use data on three teacher residency programs across the country—Res Ed, City Teach, and Teacher Prep[1]—to explore whether hosting a resident in the classroom improves student learning, as measured by improved TESes, relative to teachers who do not host residents in their classrooms. **We find some evidence that having a resident in the classroom can improve student learning, with TESes being higher for host teachers at two programs.** We do not find a positive effect for the third program. We discuss the implications of our findings in the body of the report. It should, however, be noted that we should not compare the impact of hosting residents across programs because each program defines and measures the outcome variable (i.e., TES) differently.

DS programs are direct service programs that aim to influence student outcomes by facilitating alternative staffing arrangements within schools. To achieve this, they train young cohorts of individuals who ultimately provide a range of support services to students. We explore whether, by virtue of the fact that staff members receive hands-on experience in schools, they go on to be more effective teachers. We use student- and teacher-level data to explore whether alumni of three DS programs— Teach Plus, Ed Serve, and Urban Corps—are more effective than a comparison group of non-alumni teachers. **Using student academic outcomes as a proxy of teacher effectiveness, we find no evidence that DS alumni are more effective than a sample of matched comparison teachers. We do find some evidence that they have higher retention rates.** Below we summarize findings related to each research question addressed in the report.

[1] This report uses pseudonyms throughout to conceal the names of both residency and differential staffing programs.

Key Findings from Study

Residency programs	
Research Questions	**Findings**
Primary question	
What is the added value of hosting residency personnel in a classroom, as measured by TESes, a measure that captures students' academic performance?	In two programs that follow a very similar residency model—Res Ed and City Teach—host teachers with a resident in classroom have higher TESes compared to similar teachers who do not host residents. In Teacher Prep, which follows a different residency model, we see no significant difference for host teachers compared to comparison teachers.
Supplementary questions	
How do host teachers and residents describe their co-teaching relationship and joint activities within the classroom?	Host teachers and residents both report frequent (defined as more than monthly) participation in joint activities with each other. Both groups report engaging in the following activities commonly: examining strategies for classroom management and effective instruction; examining student progress and how to adapt their teaching approach to meet students' learning needs; and having dedicated meeting time. The residents' responses also indicate that they generally feel supported by the host teachers.
Are positive co-teaching relationships positively associated with student gains?	Using aggregate school-level data on residents and host teachers from Res Ed and City Teach, we generally do not find a significant association between the quality of the co-teaching relationship and students' proficiency in math and English Language Arts (ELA).
Do schools with a larger presence of residents and resident alumni show larger gains in student performance?	Using aggregate school-level data on residents from Res Ed and City Teach, we find no significant association between number of residents in a school and the school's proficiency rate in either math or ELA.
DS programs	
Research Questions	Findings
Primary question	
Are teachers who worked with DS programs prior to becoming teachers more effective (as measured by student academic outcomes) than teachers who took more traditional routes to becoming teachers?	Using students' state test scores as a proxy of teacher effectiveness, we find no evidence that DS alumni teachers are more effective than a matched sample of comparison teachers. The results do not change when we conduct the analysis on the pooled data for all teachers and subjects together.
Supplementary questions	
What proportion of DS front line staff members go on to become teachers?	Based on alumni who responded to a survey, approximately 30 percent of DS alumni become teachers. Teach Plus has the highest percentage of alumni who became teachers (36 percent), followed by Urban Corps (~26 percent) and Ed Serve (25 percent).

Do teachers who worked with DS programs prior to becoming teachers have higher retention rates than teachers who took more traditional routes to becoming teachers?	Across the three cohorts of teachers included in the analysis, DS teachers have higher retention rates compared to the matched comparison group of teachers, with the difference being statistically significant. In the pooled sample, the retention rate of DS alumni teachers is higher than that of the comparison teachers by almost six percentage points, and the difference is highly statistically significant. Ed Serve alumni have a slightly higher retention rate (at 100 percent) compared to Teach Plus alumni (at 97 percent).
Is the population of teachers who come from DS programs more racially/ethnically diverse than the population of teachers who took more traditional routes to becoming teachers?	Teachers who come from DS programs are significantly more likely to be Asian but significantly less likely to be Latinx. This pattern was primarily driven by Teach Plus.

INTRODUCTION

High-need school districts partner with a variety of external service providers to address specific student and staffing needs that they cannot address on their own due to human and/or financial resource constraints. These external service providers are often (though not necessarily) non-profit organizations that offer niche programs to districts. By way of program design, non-profit partners can target students, teachers, or district infrastructure with their interventions. Programs that are designed to offer direct benefits to students are often called "direct service" programs. As an illustration, a school or district may partner with a college access program to assist students with the college application process; or they may partner with a tutoring program to offer high quality tutoring to some or all students in a school. Other partner programs offer benefits to teachers in the form of professional development (PD). And still other programs offer districts innovative ways to systematically source and recruit effective teachers over the long term. For instance, local colleges or graduate schools may partner with a school district to place teachers in training in classrooms as part of their training; school districts in turn benefit by gaining access to new cohorts of teachers. These local training institutions act as a teacher pipeline, particularly in high need districts that may face challenges with recruitment.

Each type of partner program presumably offers a narrow set of direct benefits to school districts. Direct service programs aim to impact student outcomes in one way or another. Teacher PD programs aim to improve teacher pedagogy. And teacher pipeline programs help districts source talent. To date, research has focused on studying the direct benefits a program aims to provide. For example, researchers study the impact of tutoring programs on student achievement in high need schools. They study the benefits of PD programs on teacher practice. And they study how effective teacher pipeline programs are at training, placing, and retaining high quality teachers.

However, because these programs operate in broader school ecosystems, each type of program may also have ancillary or indirect benefits. A program that strives to impact teachers may also impact students. Likewise, a program that aims to directly serve students may also have benefits for program staff and teachers. This report aims to study the ancillary benefits of two types of programs: one is a specific type of teacher training model called *residency programs* and the other is a type of direct service model called *differential staffing (DS) programs*.

Residency programs provide experiential learning opportunities to teachers in training by placing them in classrooms alongside host teachers for a full academic year. They are typically administered by graduate schools, which partner with local school districts with the agreement that residents can train in their classrooms in exchange for a commitment that the resident will teach in the partner district (typically for a few years) upon graduation. By acting as a teacher pipeline, residency programs meet a specific staffing need for districts. However, they also place residents in classrooms for a full year, which has multiple potential ancillary benefits. First, providing lead teachers with a co-teacher lowers the student-teacher ratio in classrooms. Second, having an additional adult in the classroom gives students another adult with whom they can forge a relationship. Third, having to host and possibly coach or mentor a resident may lead to better teaching practices among host teachers. In this way, it is theoretically plausible that having a resident in the classroom could benefit students in classrooms with both host teachers and residents.

DS programs provide alternative staffing models to schools. These programs place staff members in schools to provide various types of services, ranging from tutoring to classroom support to after school assistance for struggling students. DS programs generally aim to improve student outcomes in high-need schools. To do so, they usually deploy cadres of young people (e.g., recent college graduates) to provide specific services in the schools. These young people, in turn, gain hands-on experience working in school settings and interacting with students. This experience may shape their career aspirations by persuading them to pursue teaching or, on the flipside, dissuading them from doing so. In this way, DS programs may indirectly serve as teacher pipeline programs by recruiting, training, and deploying generations of future teachers.

This report seeks to explore the ancillary benefits of residency and DS programs by asking the following questions:

- What is the added value of hosting residency personnel in a classroom, as measured by improved TESes during the year of the residency?
- Are teachers who worked with DS programs prior to becoming teachers more effective (as measured by student academic outcomes) than teachers who took more traditional routes to becoming teachers?

By answering these questions, we aim to move away from understanding school partner programs as providing niche services to re-framing their benefits within the broader ecological system of a school district, where residents routinely interact with and potentially affect students, and staff from DS programs are exposed to and shaped by the many highlights and challenges of teaching in high-need schools.

In addition to the above questions, we also explore several related questions:

- Residency programs: What are the ways in which residents and their host teachers interact and collaborate within a classroom? Is there an association between strong co-teaching relationships between residents and host teachers and student performance? Finally, do schools with a larger number of residents demonstrate greater student gains?
- DS programs: What percentage of DS staff members go on to become teachers? Do DS alumni have higher retention rates than teachers who did not previously work for DS programs? Are cohorts of DS alumni teachers more diverse than teachers who did not previously work for DS programs?

The report focuses on programs' impacts on outcomes that are secondary to their primary outcome goals, which, in the case of residency programs, includes training future teachers and, in the case of DS programs, includes improving student outcomes. The findings presented herein do not reflect the programs' effectiveness in terms of meeting their primary outcome goals. Thus, consumers of this report should not use the report to make judgements about the effectiveness of these programs at achieving their primary goals.

This report is divided into two sections. We begin with residency programs, offering a description of the residency program partners included in our study, our data and methodology, and results. We then turn to DS programs, where we do the same. We conclude with a summary and discussion.

RESIDENCY PROGRAMS

Teacher residency programs are designed to provide pre-professional, experiential learning opportunities for teachers in training (i.e., residents). Modeled after medical residencies, residents participate in a year-long apprenticeship where they are paired with a classroom teacher and work in that teacher's classroom during the school year. Residents typically begin the year observing, co-teaching, and in some programs, being coached by the host teacher; as the year goes on, they gradually become the lead teacher in the classroom. To combine practice with theory, residents simultaneously take graduate coursework to earn a master's degree. Residency programs typically provide a stipend to residents as they learn to teach in the classroom.

Some residency programs use what is called a mentor model. The National Center for Teacher Residencies (NCTR) advocates for this model, wherein host teachers are selected to serve as mentors based on their experience, efficacy in the classroom, and willingness to act as a mentor and serve as a model of success in a high-needs classroom.[2] Under this model, host teachers provide mentorship to the residents over the course of the year, offering guidance and support, as well as structured feedback based on their observations. Residency programs typically provide training to mentors to support this work, along with a stipend in exchange for their efforts.

Other residency programs use an alternative approach, called the host model, where teachers are selected primarily based on their willingness to host a resident. In this model, host teachers do not coach or mentor residents. Instead, the residency program provides necessary coaching. The reason for this model is that it is difficult to find and train teachers to act in the mentor role, so programs instead look for teachers who are willing to host residents and then provide mentorship and support themselves. These teachers also typically receive a small stipend.

Residency programs generally serve under-resourced school districts, where teacher recruitment, training, and retention prove to be challenging. They aim to recruit diverse cohorts of residents and, upon graduation from the program, often require the residents to serve in the same school districts (where they were trained) for at least a few years. In this sense, they offer school districts a pipeline by which they can recruit and retain talented teachers who will be ready to serve students as soon as they earn their teaching certificates. Residency programs therefore often see themselves as being better positioned to meet school districts' needs than traditional teacher training programs.

As described in the "Introduction" section, this report explores an ancillary benefit of residency programs. Specifically, we are interested in exploring whether hosting a resident in the classroom has positive benefits on student learning in that classroom. There are multiple mechanisms through which hosting a resident could promote student learning. First, having an additional adult in the classroom lowers the student-teacher ratio, enabling teachers to more easily facilitate small group learning and help provide individualized instruction to students. In a typical classroom, there is considerable variation in students' abilities and learning needs and a single teacher has limited capacity to address this heterogeneity. Having an additional adult in the classroom helps teachers differentiate their instruction and support, which in turn helps students learn content more quickly. Second, having an additional adult in the classroom gives students another adult with whom they can forge a relationship. Students who do not identify closely with the lead teacher may find that they connect

[2] National Center for Teacher Residencies, 2019. *About*. Retrieved from https://nctresidencies.org/about/residency-model-teacher-mentor-programs/.

with the resident, which bolsters their attachment to school overall and can promote learning. Third, having to host and possibly coach or mentor a resident may lead to better teaching practices among host teachers. They may reflect on their own practice more regularly, want to model strong teaching practices in the classroom, and find themselves planning in advance in order to better support and utilize the resident. All of these actions would presumably have positive benefits for student learning.

That said, it is plausible that having a resident in the classroom could potentially lead to lower student achievement than in classrooms without residents. If the lead teacher is overwhelmed or distracted by the presence of another adult in the classroom, does not know how to leverage the extra help, or does not have a healthy relationship with the resident, then the resident's presence in the classroom could ultimately be counterproductive to student learning.

Below we draw on data from residency programs and their partners to explore the relationship between hosting a resident in the classroom and student achievement. Our primary research question related to residency programs is:

- What is the added value of hosting residency personnel in a classroom, as measured by TES during the year of the residency?

In addition, we ask the following related questions:

- What are the ways in which residents and their host teachers interact and collaborate within a classroom?
- Is there an association between strong co-teaching relationships between residents and host teachers and student performance?
- Do schools with a larger number of residents demonstrate greater student gains?

To answer our research questions, we draw on data from three residency programs. Two programs – Res Ed, and City Teach – use the mentor model and the third program – Teacher Prep – uses the "host teacher" model. For simplicity, in this paper, we refer to all teachers who host a resident in their classroom as "host teachers," irrespective of whether they use the mentor model.

PROGRAM BACKGROUNDS

We began by gathering exploratory information from three residency programs Overdeck recommended to us: Res Ed, Teacher Prep, and one additional program. We also spoke with experts at Public Impact and NCTR, as well as other experts in the residency space (David Rosenberg at Education Resource Strategies and Dr. Cynthia Osborne, who had conducted previous research on residency programs in Texas) to learn more about residency programs and the overall landscape. This exploratory work gave us a good sense of existing research on residency programs and jump started our thinking regarding questions we should ask prospective residency program partners in this research project.

Overdeck initially identified one criterion on which they wanted to base the selection of residency program partners: specifically, that the program operated nationally or desired to scale nationally. Additionally, we identified a second criterion for inclusion in the study: that the program model pair residents with one host teacher at a time for a sustained period of time. Although all three programs we talked to met the first criterion, only two of the programs, specifically Res Ed and Teacher Prep, kept residents with one host

teacher in their classroom throughout the day. The third program recommended by Overdeck employed a different residency model, where residents moved around throughout the day and did not stay in the same classroom. Furthermore, even if the residents were in the same room, their function sometimes varied from being an observer to a tutor to teaching the entire class, depending on the time of year and the needs of the school. We, therefore, decided to exclude this program from the study and partner with Res Ed and Teacher Prep.

During our conversation with NCTR, which has partnered with approximately 35 residency programs to date, we came to learn about City Teach, an NCTR partner program that uses a model similar to Res Ed. However, it is a regional program with no intention to scale nationally. That said, though City Teach did not meet the first criterion, there were other reasons that made City Teach a good fit as a potential partner. They had a sizeable number of potential residents we could include in our study sample, and they were one of NCTR's partner organizations. This meant we could also utilize NCTR survey data from these two programs to try to answer the supplementary research questions on residency programs. We, therefore, added City Teach as our third residency program partner for our study.

Below we describe how each of the program partners selects its mentor teachers. To protect the anonymity of the organizations that participated in this study, we provide limited information on the background of the organization and their program model.

Res Ed: With guidance from Res Ed, principals at partner schools nominate high-quality classroom teachers to serve as host teachers to residents. Principals recommend teachers based on a host of criteria, including overall effectiveness with students, a demonstrated ability to collaborate with colleagues, and affirmative surveys from students, parents, and peers. An additional important characteristic of potential host teachers is that they must be adept at making their teaching practice transparent and articulating their thinking to residents. The principal also takes into consideration potential host teachers' willingness to share their classroom for an entire year with a resident. Res Ed seeks to source host teachers on these multiple dimensions not only to identify quality host teachers but also facilitate a good match between host teachers and residents. Depending on the host teacher pool at a partner school in a given year, it is possible Res Ed will not place residents in a partner school if they do not feel it will be a good fit for the residency cohort.

City Teach: Teachers complete an application process to be considered a host teacher. A teacher must have at least three years of teaching experience in one of the district's high-needs school, commit to hosting a resident for a full academic year, and obtain the approval of their principal to apply. Additionally, candidates must hold a current license in their subject area and meet district requirements for student teaching supervisors. They are also expected to obtain a certain level of effectiveness, as measured by a district-specific teacher effectiveness measure. In addition to the application, City Teach staff observe and interview prospective host teachers. The observation and interview—coupled with data on the teacher's commitment to student achievement through planning, instruction, classroom management, and evaluation of self and students—form the basis for host teacher selection.

Teacher Prep: As mentioned earlier, one major difference between Teacher Prep and the other residency programs included in our study is that Teacher Prep refers to their host teachers as "hosts" as opposed to "mentors," as in the case of the other two programs. The intentional language reflects Teacher Prep's belief that it is difficult to control the pipeline of high-quality host teachers. As such, they are really looking for a teacher to "host" a resident for a year, and Teacher Prep will take responsibility for coaching the resident.

The selection of host teachers relies heavily on principal recommendation and willingness on the part of the host teacher to grow their own practice and to share their classroom with another person. Furthermore, Teacher Prep provides limited training to host teachers. The training is focused on the host teacher's role with respect to the resident. Teacher Prep refrains from calling this "professional development" for the host teacher. This limited training is consistent with the expectation that the host teachers in this model are not taking on the responsibility of developing the resident while they are in the classroom. Instead, residents are assigned their own coaches, as mentioned above, and offered other opportunities for professional enrichment.

DATA AND STUDY SAMPLE

To answer the primary research question, we use data from the three residency programs cited above: Res Ed, City Teach, and Teacher Prep and their respective school partners (described below). With customized memoranda of understanding (MOUs) in place, we requested and received deidentified, individual-level data on all teachers, students and schools from each residency program's partner school districts or charter management organizations (CMOs). For each program, we also requested and received rosters of residents placed in schools, along with a link file to connect residents to their host teachers. These partner districts were selected based on a range of factors, such as willingness and ability to share data and the depth and length of their partnerships with their respective residency programs. Below we discuss the study sample for each residency program:

Res Ed: Our study focuses on residents placed with host teachers within schools at one of Res Ed's partner CMOs—Charter Plus Academy (Charter Plus). We selected Charter Plus to participate in the study because Res Ed has partnered with Charter Plus for several years and, therefore, would be able to offer several years of data for analysis. Table 1 summarizes the years for which Charter Plus (and the other respective partner districts) provided data and that are included in the study. For Res Ed we received data on five cohorts of residents placed with host teachers from 2012–13 to 2016–17.

Table 1. Academic years of data included in the study, by program. An X indicates data was provided by the program. Source: Program records.

	12–13	13–14	14–15	15–16	16–17	17–18
Res Ed	X	X	X	X	X	
City Teach		X	X	X	X	X
Teacher Prep					X	X

After excluding host teachers with missing data in either baseline year(s) or the outcome year, the sample includes 187 teachers across the five cohorts. Table 2 provides a breakdown of the number of teachers included in the sample by year. In our sample, most host teachers (80 percent) serve as a host teacher for a single year, with only 20 percent serving as a host for more than a year. Furthermore, within each year, each host teacher is assigned only one resident and, likewise, each resident is assigned to only one host teacher. Therefore, there are only unique host teacher and resident combinations for each year. Since teachers may work with residents in more than one year, there are 148 unique host teachers included in the sample.

Table 2. Number of teachers by program and year of service. Source: Program academic datasets.

Academic Year	No. hosts	Pct. serving as hosts out of unique hosts	No. teachers (all)	Pct. teaching out of unique teachers (all)
Res Ed				
12–13 only	22	15%	68	7%
13–14 only	13	9%	75	8%
14–15 only	14	9%	48	5%
15–16 only	31	21%	79	8%
16–17 only	39	26%	179	19%
Multiple years	29	20%	497	53%
Unique no. teachers	148	100%	946	100%
Total no. with teachers repeating across years	187		1894	
City Teach				
13–14 only	11	13%	468	13%
14–15 only	17	20%	248	7%
15–16 only	7	8%	101	3%
16–17 only	10	12%	205	6%
17–18 only	15	18%	426	12%
Multiple years	24	29%	2088	59%
Unique no. teachers	84	100%	3536	100%
Total no. with teachers repeating across years	113		7403	
Teacher Prep				
16–17 only	39	42%	149	22%
17–18 only	44	48%	160	23%
Both years	9	10%	382	55%
Unique no. teachers	92	100%	691	100%
Total no. with teachers repeating across years	101		1073	

City Teach: City Teach places its residents in both public and charter schools in two school districts: Regional School District (RSD) and a local turnaround district. Our study includes residents placed with host teachers in both public and charter schools within the RSD district. We exclude teachers in the turnaround schools because we were unable to request data from this district. City Teach reports that, since its inception, approximately 84 percent of its residents have been placed in RSD, divided between charter schools (18 percent) and traditional public schools (66 percent). As such, we did not perceive lack of access to data from the local turnaround district as a limitation of the study.

RSD provided data on five cohorts of residents serving in schools between 2013–14 and 2017–18. After excluding cases that are missing data for either baseline year(s) or the outcome year, the sample includes 113 teachers across the five cohorts. (Table 2 shows the breakdown by year.) In our sample, most host teachers (71 percent) serve as a host for a single year, with only 29 percent serving for more than one year. As with Res Ed, within each year, each host teacher is assigned only one resident and each resident is assigned to only one host teacher. There are some rare cases where a resident could be assigned to two host teachers.[3] In these cases, we use the teacher-resident combination for the teacher with whom the resident spent most of the year. Since teachers may work with residents in more than one year, there are 84 unique teachers included in the sample.

Teacher Prep: This study draws on data from one of Teacher Prep's partner CMOs, the Relate Network (Relate). Relate provided data for the 2016–17 and 2017–18 cohorts of residents. After excluding cases that are missing data for baseline or outcome years, the sample includes 101 teachers across the two cohorts. (Table 2 includes the breakdown by year.) As with the other two programs, most host teachers (90 percent) serve as hosts for a single year, and only 10 percent serve as a host for both years. Within each year, each host teacher could be assigned multiple residents (in the overall sample, 11 percent were assigned two residents while 89 percent were assigned only one). Likewise, each resident could be assigned to multiple host teachers (56 percent were assigned to two host teachers, spending one semester with each, while 43 percent were assigned to only one host teacher over the entire year[4]). Since teachers may work with residents in more than one year, there are 92 unique teachers included in the sample.

METHODOLOGY

Our primary outcome measure is the TES (defined below under "Dependent variable"). We use a consistent methodology to conduct our analysis across all three programs.[5] Specifically, we run a multilevel regression model to examine whether the TES in the outcome year is significantly different for teachers who host residents compared to those who do not, after taking into account the TES in prior years (i.e., TES baseline), years of teaching experience, and teachers' school characteristics (e.g., percentage students who are English language learners (ELLs), have individualized education programs (IEPs), are economically disadvantaged, are Latinx or black, Asian or white). The model includes a binary variable indicating whether the teacher serves as

[3] According to City Teach, this only happens when there is an extenuating circumstance that leads to a mid-year change of mentor (e.g., mentor's medical leave, mentor-resident relationship not working, etc.).
[4] These estimates are not shown in Table 2, though additional information is available upon request.
[5] With the exception of Teacher Prep, we do not have data for pre-residency years (i.e., before the inception of the residency program) to be able to match host teachers to teachers not hosting residents on their trend in TESes from pre-residency year(s).

a host teacher (1 if host teacher, and 0 if not). The coefficient on this variable estimates the mean difference between host teachers and comparison teachers on the outcome measure.

We run the analyses separately by program and by year. For each program and year, the comparison group includes teachers in the same partner districts/networks who did not work with a resident in that year. It is possible that working with a resident has an enduring impact on a teacher's practice, such that s/he may be a more effective teacher after working with a resident, even if s/he never again works with a resident. We therefore drop teachers from the dataset in all years succeeding the year they host a resident if they do not host resident in the subsequent year/s so they are not included in the comparison group. Thus, teachers who host a resident are only included in the dataset for the years in which they host a resident.

Statistical model

Our primary multilevel regression model (i.e., specification 1) is as follows:

$$\text{TESoutcome}_{ijt} = \alpha_0 + \beta_0\,\text{Host}_{ijt} + \phi_0\,\text{TESbaseline}_{ijt} + \lambda_0\,\text{TE}_{ijt} + \alpha_1\,(\text{School})_{it} + \beta_1\,\text{year} + \varepsilon_i + \epsilon_j,$$

where, TESoutcome_{ijt} is the TES for teacher *i* in school *j* in year *t*. Host_{ijt} is a binary variable that captures whether teacher *i* in school *j* in year *t* is a host teacher (= 1 if host teacher; 0 if comparison teacher). TESbaseline_{ijt} is the baseline TES for teacher *i* in school *j* in year *t*, where baseline TES is calculated by averaging the scores from all available baseline years. TE_{ijt} captures number of years of teaching experience for teacher *i* in school *j* in year *t*. $(\text{School})_{it}$ represents aggregate school-level variables such as percentage ELL and IEP that vary by school and by year. Variable year represents year fixed effects that we account for in the model since we have data for multiple years. Finally, $\boldsymbol{\varepsilon}\mathbf{i}$ captures the random error associated with teacher *i* and ϵj captures the random error associated with school *j*. We use a mixed effects model to account for multilevel clustering at the teacher (since there are multiple scores across time for each teacher) and school levels (as there are multiple teachers from the same school), which allows us to report correct standard errors for the impact estimates. The model therefore accounts for these clustering so that the constant coefficient is allowed to vary across schools and teachers.

In addition to our primary model (specification 1), we run two additional models as checks on the robustness of our findings. First, some host teachers are in the dataset for multiple years because they host residents for multiple years. It is possible that these teachers are particularly adept at both teaching and leveraging the skills of their residents, and that their presence in the dataset multiple times is driving the positive effect. Therefore, in our second model specification (i.e., specification 2), we only keep host teachers in the sample in the *first year* they serve as a host. Otherwise, the model is identical to specification 1. Moreover, since TESes (particularly in the case of CT), typically account for the characteristics of the schools in which teachers teach, by including them in our models, we are essentially adjusting for them twice. Thus, in our third specification (i.e., specification 3), we continue to control for teacher-level characteristics but drop school-level characteristics from the model.

Dependent variable

The outcome variable is the TES in the outcome year (i.e., the year of analysis). Across all of the districts and CMOs included in the study, each teacher receives a TES each academic year that is considered a measure of his/her performance in that year. Although each program defines and measures the TES differently, student performance is always a primary factor in the TES. Below we discuss how each district or CMO defines its TES.

Res Ed: Charter Prep's TES is a continuous variable that aligns with five effectiveness categories (see Table 3). Charter Prep's TES accounts for student growth (i.e., student growth percentile using a student growth percentile model) as well as scores based on principal evaluations, student and family surveys, and teammate surveys. In the years included in this study, approximately 40 percent of the TES is based on student achievement growth, with 30 percent based on growth of each teacher's students and 10 percent based on growth in school-wide achievement. (For teachers in non-tested grades and subjects, student achievement growth is based only on schoolwide growth.) Approximately 40 percent is based on observed teacher practice, 10 percent on student survey feedback, five percent on parent survey feedback, and five percent on peer survey feedback. The scores can be converted into an ordinal variable with five levels (see Table 3). On Charter Prep's scale, most teachers are rated either effective (28.4 percent) or highly effective (62.2 percent).

Table 3. Res Ed teacher effectiveness score description and frequencies pooled across all treatment years. Source: Charter Prep's records.

Teacher Effectiveness Category	Frequency (%)
Entering (1)	0.0
Emerging (2)	1.6
Effective (3)	28.4
Highly Effective (4)	62.2
Master (5)	7.8

City Teach: The TES for City Teach is a value-added score (VAS) based on value-added modeling, a method of teacher evaluation that measures the teacher's contribution in a given year by comparing the current test scores of their students to the scores of the same students in previous school years. The state estimates teacher VAS. RSD provided scores for each teacher who taught in the RSD district in the years included in the study. The scores are calculated by assessing the performance of a teacher's students on the end-of-year, state-mandated assessments, taking into account the students' past performance on such assessments. The score therefore captures the degree of growth a teacher's students exhibited on achievement tests from one year to the next. When students outperform their past performance, it will raise the teacher's score, and vice versa.

Teachers who taught students in tested subjects receive a separate VAS effectiveness score for each subject and for each class they taught. For the purpose of our analysis, we create a composite VAS score for each teacher for each year, by using a weighted average based on the number of students the teacher taught in each tested subject and class. For example, if a teacher taught geometry to 20 students and algebra to 60 students, the calculation gives the algebra score three times more weight than the geometry score. This approach is consistent with the approach used by district researchers in their analyses.

Each teacher's annual value-added composite score was calculated using the value-add index variable (a standardized continuous variable), which is calculated using the student growth estimate (produced by the state's statistical model) divided by its standard error. The values of the index can be interpreted in terms of effectiveness levels of one to five (an ordinal variable with which most state educators are familiar), as indicated in Table 4. Under this rating scheme, most teachers are considered to have average effectiveness.

Table 4. City Teach value-add levels, label description based on value-add index, and frequencies pooled across all treatment years. Source: RSD records.

Label (Performance level)	Underlying Index	Frequency (%)
Least effective (1)	Index < -2	8.1
Approaching average effectiveness (2)	-2 ≤ Index < -1	8.8
Average effectiveness (3)	-1 ≤ Index < 1	65.7
Above average effectiveness (4)	1 ≤ Index < 2	8.8
Most effective (5)	2 ≤ Index	8.5

Teacher Prep: The TES for Teacher Prep is an "overall performance measure" of teachers based on Relate's teacher evaluation system, which takes a multi-measure approach that reflects four goals for all teachers from kindergarten to grade 12. Two of the goals are "outcome" based and are tied to student performance on assessments; the other two are "effort" based and are tied to classroom observations and an end-of-year evaluation. Appendix A explains how the points from each goal are then used to determine the overall performance calculations. The overall performance value or what we refer to as the TES is an ordinal measure of teacher performance that can range from one ("working toward high bar") to three ("exceeding high bar") in the treatment years. Table 5 shows the frequency of each level among the teachers in our sample. Nearly two-thirds of teachers receive a "meeting high bar" rating.

Table 5. Teacher Prep teacher effectiveness score overall performance measure description and frequencies pooled across all treatment years. Source: Relate's records.

Overall Performance Description	Overall Performance Level	Frequency (%)
Working toward high bar	1	14.1
Meeting high bar	2	65.5
Exceeding high bar	3	20.4

Independent variables

Teacher level:

- **Host teacher:** This is a binary variable capturing whether a teacher is a host teacher in any given year (1 if host teacher, and 0 if not). Each program provided a roster indicating to which teacher(s) each resident had been assigned in each academic year. Using this roster, we create a binary indicator variable that indicates whether a teacher hosted a resident in a given year. The coefficient of this variable gives the mean difference between host teachers and comparison teachers on the outcome measure.
- **Baseline TES**: For each program, we utilize all available data on teachers' prior years' TESes to estimate an average baseline TES for each treatment year included in our analysis. Table 6 shows

how we define average baseline TES for each program, and the corresponding outcome year TES for each year, with the latter being our primary outcome variable.

- **Teaching experience (in years):** Teacher experience is an indicator of teacher effectiveness and student performance. We therefore control for number of years of experience in our model, except in the case of City Teach where data are not available.

Table 6. Academic years included in baseline TES estimates for each cohort, by program.*

	11–12	12–13	13–14	14–15	15–16	16–17	17–18
Res Ed							
Avg. Baseline TES	U V W X Y Z	V W X Y Z	W X Y Z	X Y Z	Y Z	Z	
Outcome TES		U	V	W	X	Y	Z
City Teach							
Avg. Baseline TES		V W X Y Z	W X Y Z	X Y Z	Y Z	Z	
Outcome TES			V	W	X	Y	Z
Teacher Prep							
Avg. Baseline TES			YZ	YZ	YZ	Z	
Outcome TES						Y	Z

* Cohorts are indicated by letter and color.

School level:

- In addition to prior experience, a teacher's TES might differ by the characteristics of the schools they serve. We therefore control for school-level characteristics including percentage of students who are special education, percentage of students who are ELLs, percentage of students who qualify for free or reduced-price lunch or have an economically disadvantaged status, and percentage of students who are black or Latinx versus white or Asian.

FINDINGS

Descriptive statistics

Table 7 compares characteristics of host teachers to teachers who did not host residents. For each residency program, we show average characteristics for both groups of teachers using the pooled data for all years (for comparisons of characteristics for each year, see Appendix B). The average TES in baseline year/s is higher among Res Ed and City Teach host teachers relative to comparison teachers; however, among Teacher Prep teachers, the TES for host teachers is comparable to that of other teachers. Res Ed host teachers have roughly 1.3 additional years of teaching experience compared to other teachers; the opposite is true among Teacher Prep host teachers (we do not have teacher experience data for City Teach). Notably, this is what we would have expected, given the different host teacher recruitment models used by Res Ed versus Teacher Prep (described earlier in this report). When we compare the school-level characteristics for the host teachers versus other teachers, the host teachers at all three programs seem to serve in schools with a higher proportion of economically disadvantaged students: the percentages that are economically disadvantaged and Latinx or black are higher in schools where host teachers serve and the percentage of white and Asian students is a bit lower. There is not much difference between the characteristics of schools at which the two

groups teach based on school-level characteristics such as percentage special education (SPED) and limited English proficiency (LEP).

Table 7. Characteristics of host teachers compared to teachers who did not serve as hosts. Source: Residency and partner district administrative records.

	Res Ed		City Teach		Teacher Prep	
	Host	Other	Host	Other	Host	Other
Teacher characteristics						
Avg. TES in baseline year/s	3.6	3.2	1.0	0.4	2.2	2.2
Years teaching experience	6.6	5.3	--	--	3.4	4.5
School level characteristics						
Pct. SPED	1.3	1.3	12.0	11.4	7.1	6.6
Pct. LEP	22.7	23.8	--	--	35.4	31.4
Pct. Free and reduced-price lunch	80.5	78.3	89.0	79.4	83.0	72.3
Pct. Latinx and black	87.9	84.6	91.8	87.7	94.1	82.7
Pct. white and Asian	9.0	11.6	7.6	11.3	4.4	15.2
N (teachers)	187	1707	113	7290	101	972

Note: -- indicates that program did not provide data for this measure.

Regression results

Table 8 summarizes the results of the analysis using the three model specifications, as well as results broken out by year using specification 1. (Appendix C provides detailed regression results using specification 1.) The second, third, and fourth columns summarize results on all years of data, using the three different specifications. Subsequent columns of Table 8 present results for each year using specification 1.

Res Ed: The results for all years using specification 1 indicate that, controlling for baseline TES and other teacher- and school-level characteristics, Res Ed hosts had an average outcome TES that was .11 points higher than the average TES of teachers who did not host residents (3.77 versus 3.67). This difference is statistically significant ($p<.01$). Although both groups' mean TES falls closer to the teacher effectiveness category of "highly effective," being an Res Ed host is associated with an increase in one fifth of the standard deviation of outcome TES (the outcome TES has a SD of 0.56 and $0.11/.56 = 0.2$).

Figure 1 shows the results from specification 1 visually. The graph was created using binscatter (a command in Stata software package), which grouped the TESbaseline variable into equal-sized bins, and then computed the mean of the TESbaseline variable and TESoutcome variables within each bin, to create a scatterplot of these data points.[6] Each dot shows the average TESoutcome for a given level of TESbaseline for Res Ed hosts (in red) and comparison teachers (in blue). Finally, binscatter plotted the best linear fit line, from a multivariate regression of the TESoutcome variable on the TESbaseline variable. The graph therefore shows the relationship between teachers' baseline TES and their outcome TES, so that we have a general sense of what the average TES is in the outcome year, given a teacher's TES in the baseline year, for host teachers and comparison teachers separately. If host teachers indeed show more growth than comparison teachers, we would expect to see the red dots higher than the blue dots for most points along the distribution. This would

[6] Michael Stepner, BINSCATTER: Stata module to generate binned scatterplots, Statistical Software Components S457709, Boston College Department of Economics, revised 24 Nov 2013.

indicate that host teachers' TESoutcome scores were higher than those of the comparison teachers and that the effect is not attributable to just a few teachers with very high scores.

Table 8. Results from regression analyses. Source: Residency and partner district administrative records.

	All Years			12–13	13–14	14–15	15–16	16–17	17–18
Year	**Spec. 1**	**Spec. 2**	**Spec. 3**	**Spec. 1**					
Res Ed									
Host	3.77	3.77	3.76	3.74	3.64	3.85	3.77	3.87	--
Comparison	3.67	3.66	3.66	3.53	3.56	3.74	3.70	3.80	--
Diff	0.11	0.10	0.10	0.21	0.08	0.11	0.07	0.06	--
Sig?	***	***	***	***	No	No	No	No	--
N	1894	1855	1855						--
City Teach									
Host	0.65	0.83	0.81	--	0.94	1.05	0.18	0.29	0.15
Comparison	0.05	0.05	0.04	--	0.16	0.33	-0.23	-0.25	-0.02
Diff	0.60	0.78	0.78	--	0.78	0.72	0.41	0.54	0.17
Sig?	***	***	***	--	No	**	No	*	No
N	7403	7374	7374	--					
Teacher Prep									
Host	2.02	2.01	1.99	--	--	--	--	2.11	1.94
Comparison	2.04	2.04	2.03	--	--	--	--	2.15	1.95
Diff	-0.02	-0.03	-0.03	--	--	--	--	-0.04	-0.01
Sig?	No	No	No	--	--	--	--	No	No
N	1073	1064	1064	--	--	--	--		

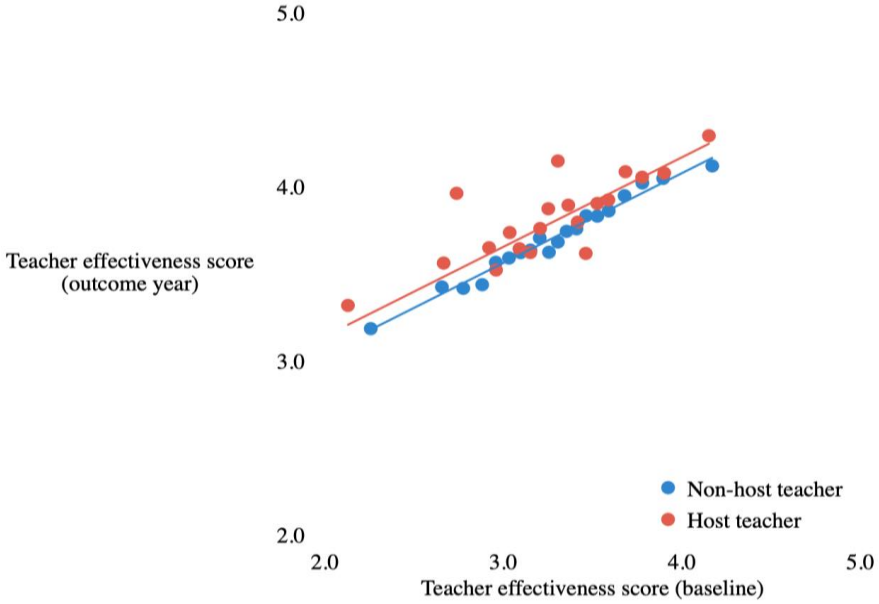
*** p<0.01, ** p<0.05, * p<0.1

In Figure 1, we see that, in most cases the red dots for host teachers are above the blue dots for comparison teachers, suggesting that the specification 1 results are robust.

As a further robustness check, in specification 2, we only keep Res Ed hosts in the sample in the first year they served as a host. The results do not change substantially under this new specification. Finally, in specification 3, we only control for teacher-level characteristics and drop school-level controls. Our results do not change much from specification 1 and are almost identical to the results derived from specification 2.

When we conduct the analysis separately by year, there is no significant difference between the outcome TES for Res Ed hosts versus other teachers in any year except for academic year 2012–13: in this year, being an Res Ed host is associated with an increase in outcome TES of 0.21, and the difference is statistically significant ($p<.01$).

Figure 1. Relationship between teacher effectiveness score (baseline) and teacher effectiveness score (outcome year) for Res Ed, based on model specification 1. Source: Residency and partner district administrative records.



City Teach: The results for all years using specification 1 indicate that, controlling for baseline TESes and other teacher- and school-level characteristics, City Teach hosts had an average outcome TES across all years (i.e., value add composite scores or index) that was 0.60 standard error units higher than that of non-hosts (0.65 versus 0.05), and the difference is statistically significant ($p<.01$).[7] Although both groups' mean composite scores fall within the Level 3 (i.e., average) performance level, being a City Teach host is associated with an increase in 0.32 of the standard deviation of outcome TES (the outcome TES has a SD of 1.9 and 0.6/1.9 = 0.32).

Figure 2a shows the City Teach results from specification 1 visually. In the graph, we see more red dots above blue dots than vice versa although it appears that there are some outliers at the lowest and highest end of the distribution that could be affecting the mean estimates. To confirm that they are not biasing the results, we remove all outliers with TESbaseline less than -2 and greater than 7. Our estimates do not change. We show the revised scatter plot in Figure 2b. In this revised chart, we can see that the host curve is more consistently over the blue curve.

In specification 2, City Teach host teachers outperformed the non-City Teach hosts by 0.78 standard error units, an increase in 0.41 of the standard deviation of outcome TES (0.78/1.9 = 0.41). In specification 3, the results do not change much from specification 2. When we conduct the analysis separately by year, there is a significant difference between the outcome TES for City Teach hosts versus other teachers in academic year 2014–15: in this year, being a City Teach host is associated with an increase in outcome TES of 0.72, and the difference is statistically significant ($p<.05$). In 2016–17, being an City Teach host is associated with an increase in outcome TES of 0.54 standard error units, but the difference is only marginally statistically significant ($p<.10$).

[7] VAS index is calculated using the growth estimate produced by the VAS statistical model divided by its standard error.

Figure 2a. Relationship between teacher effectiveness score (baseline) and teacher effectiveness score (outcome year) for City Teach, based on model specification 1. Source: Residency and partner district administrative records.

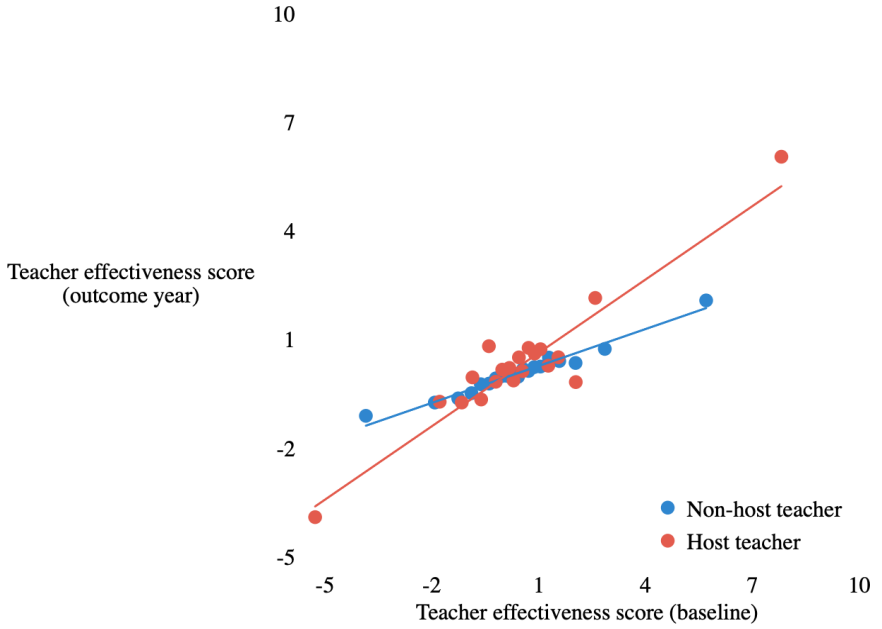
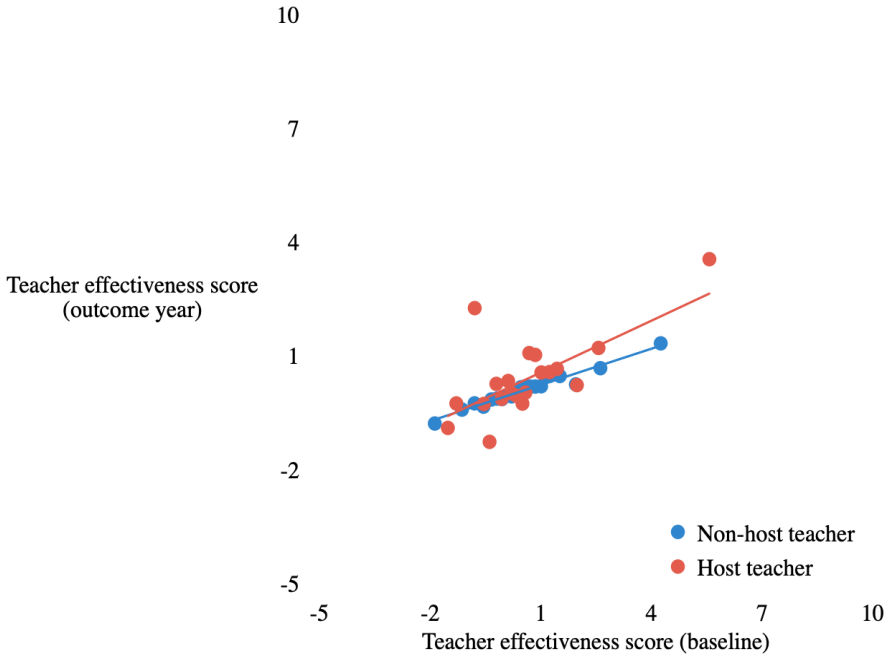


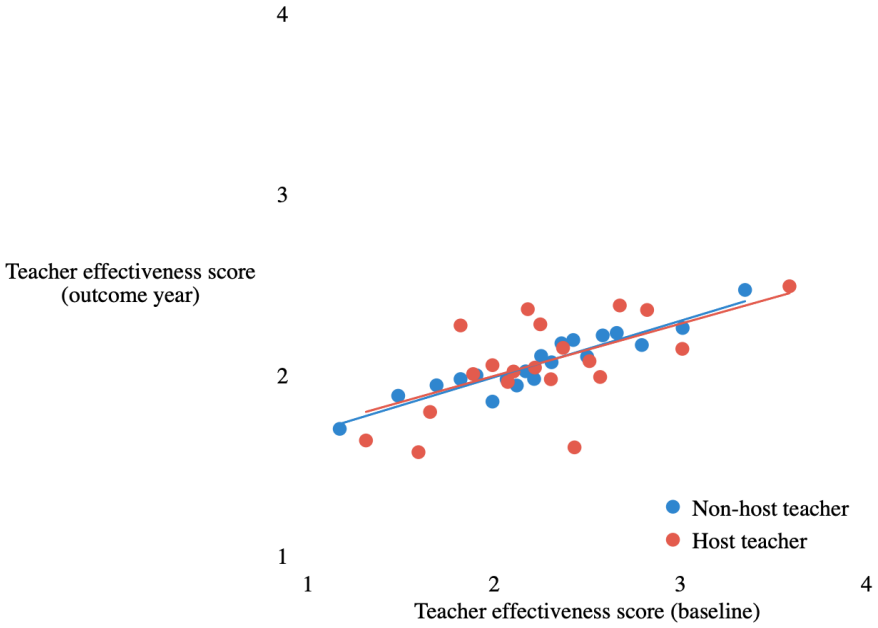
Figure 2b. Relationship between teacher effectiveness score (baseline) and teacher effectiveness score (outcome year) for City Teach, based on model specification 1, without outliers. Source: Residency and partner district administrative records.



Teacher Prep: Finally, for Teacher Prep the results for all years using specification 1 indicate that the difference between host and comparison teachers for all years was small, negative (-.02), and statistically insignificant. On average, controlling for baseline TESes and school-level characteristics, Teacher Prep hosts had an average outcome TES of 2.02, compared to 2.04 among teachers who did not host a resident. Although both groups' mean TESes fall within the second level of overall teacher performance description, being a Teacher Prep host is associated with a decrease in approximately one thirtieth of the standard deviation of outcome TES (the outcome TES has a SD of 0.59 so 0.02/0.59=0.03). The below graph shows the results from specification 1 visually. In the graph, it does not seem that any group is doing better than the other, with the red dots sometimes below the blue dots and sometimes above them.[8]

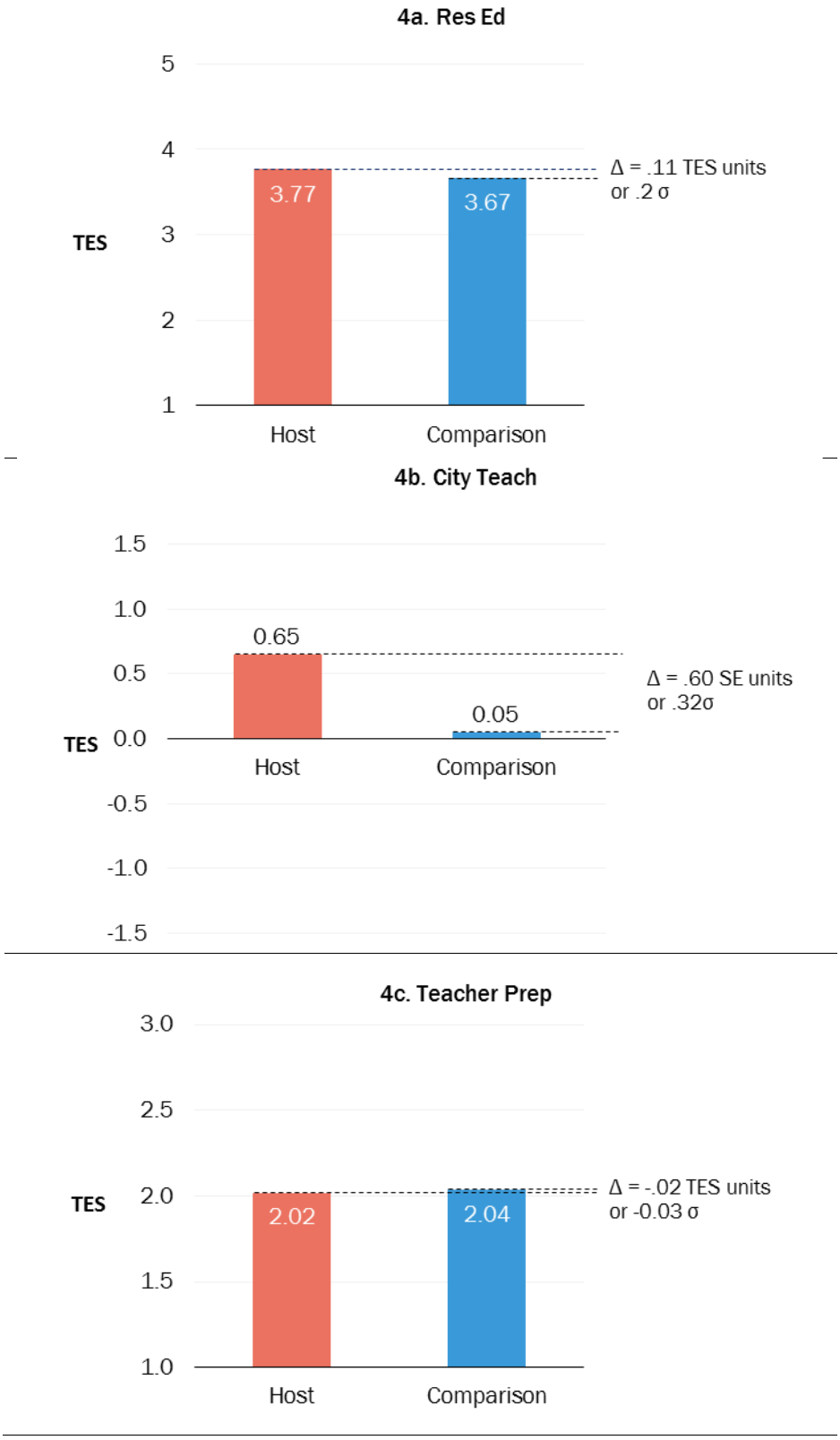
The specification 2 and 3 results do not change much from specification 1. When we conduct the analysis separately by year, there is no significant difference between the outcome TES for Teacher Prep hosts versus other teachers in either year.

Figure 3. Relationship between teacher effectiveness score (baseline) and teacher effectiveness score (outcome year) for Teacher Prep, based on model specification 1. Source: Residency and partner district administrative records.



[8] Because the Teacher Prep TES only includes three categories, as a robustness check, we also ran an ordered logistic regression to examine whether being a Teacher Prep host teacher changes the likelihood of receiving a higher rating. The results indicate there is no effect. We similarly repeated the analysis looking at the association between number of months of hosting a UT resident in the classroom and the likelihood of receiving a higher rating. Once again, there was no effect. All results are available upon request.

Figures 4a–4c. Estimated teacher TES/Standard Error (SE) in the year(s) host teachers worked with resident(s), based on specification 1 models. Source: Residency and partner district administrative records.



SUMMARY

Figures 4a through 4c summarize the effects we observed in the prior analysis. The results for Res Ed and City Teach are consistent in that we find some evidence that hosting a resident in the classroom might be associated with higher TESes of teachers, controlling for their baseline TES, teaching experience, and school-level characteristics. Since the academic performance of students in a teacher's classroom is one of the major factors (and the only factor in the case of City Teach) contributing to the calculation of a teacher's TES, higher TESes among host teachers suggest that hosting a resident in a classroom may be associated with improved student academic outcomes.

The results for Teacher Prep show no significant difference in the TESes of host teachers and other teachers. That said, we know that the residency model for Teacher Prep differs from the model used by the other two programs, which are based on the mentor model and therefore have stricter selection criteria for host teachers. Our descriptive statistics confirm that, consistent with the program model, both Res Ed and City Teach are recruiting better qualified and more experienced teachers based on average TESes and years of teaching experience. On the other hand, Teacher Prep does not emphasize teacher experience or quality when selecting host teachers. This may have implications for how well host teachers can utilize the resident in the classroom. It is possible that more experienced and more effective teachers would be better prepared to support residents. We tested this by introducing an interaction term between baseline TES and being a Teacher Prep host, as well as years of teaching experience and being a Teacher Prep host, but found no evidence of any significant difference in the TESes of host teachers based on their baseline and prior teaching experience (these results are not included in this analysis but are available upon request).

We similarly tested this for the other programs. We did not find any interaction effects in the case of Res Ed. However, we did find a significant interaction effect between being a City Teach host and the baseline TES (we did not have data on teacher experience to introduce interaction effects for experience and being a City Teach host). This suggests that the effect of being a City Teach host varies depending on the baseline TES. In other words, hosts with higher baseline TES had significantly higher outcome TES: each unit increase in baseline TES is associated with a 0.36 standard error unit increase in their outcome TES. Thus, it is plausible that host teachers are more effective not only because they get support and training from the residency programs (as in the case of City Teach and Res Ed), but also because they are more effective teachers to start with (as in the case of CT).

Furthermore, for Teacher Prep, in any given year, each host teacher could be assigned multiple residents and each resident could be assigned to multiple host teachers. Moreover, a majority of residents in our sample are paired with two host teachers within a year, thus spending on average one semester with each host teacher. This is in contrast to the other two programs where residents are typically paired with the same host teacher for an entire year. We therefore explored whether there is a significant association between length of hosting residents and TES among host teachers, finding no evidence of such (see results in last column of Table C3 of Appendix C).

The question of *why* host teachers at Res Ed and City Teach were able to translate having a resident in the classroom into higher effectiveness scores remains unanswered. As we alluded to earlier, there are two, non-mutually exclusive possibilities: having a resident in the classroom makes a teacher more effective because the training and experience improves their own practice *or* there is something about having a resident in the classroom (e.g., the lower student-teacher ratio, greater opportunity for student-teacher relationships, etc.) that improves student learning. Exploring these mechanisms in a comprehensive way is beyond the scope of

this report. However, the longitudinal nature of the dataset allowed us to explore whether the effect of having a resident in the classroom only exists in the year(s) in which a teacher is hosting a resident or if it persists after the resident leaves the classroom (i.e., in subsequent years). This may shed light on whether the effect is being driven by changes in teacher practice, which may be more permanent, or by having an additional person in the classroom. Using a mixed effects regression model similar to the model we describe above, we examined whether there was an effect among teachers who had previously hosted a resident. Because of data limitations, we limited this analysis to the Res Ed and City Teach programs. Interestingly, the results were mixed. For City Teach, teachers who previously hosted a resident were neither more nor less effective than other teachers, lending support to the hypothesis that the observed effects are being driven by actually having a resident in the classroom. Conversely, for Res Ed, teachers who had previously served as mentors were more effective, even if they no longer had a resident in the classroom. This would seem to suggest that participating as a host teacher may lead to enduring changes in a teacher's practice. (These results are all available upon request.) Although the sample sizes are small, these results are a springboard for further study on the underlying mechanisms of the effect.

Having presented and discussed our findings, we would like to caution the readers to refrain from comparing the impact of hosting residents across programs because the way the outcome variable (i.e., the TES) is defined and measured differs across programs.

SUPPLEMENTARY RESEARCH QUESTIONS

How do host teachers and residents describe their co-teaching relationship and joint activities within the classroom?

This question is motivated by a desire to better understand the extent to which residents and host teachers work together and collaborate. We explore this question for residents working with programs that use the mentor model, since it is less relevant for the host model and, as we explain below, we only have data from the programs that follow the mentor model. The ideal mentor model is one in which host teachers coach their residents in how to plan, instruct, and gauge student progress. As mentors, teachers provide both formal and informal feedback to their residents and offer structured co-planning and co-teaching opportunities. This type of coaching is not only ideal for training residents, but we would also expect student learning to be greater in settings in which the teacher and resident are synchronized and well-coordinated.

To answer this supplementary research question, we draw on stakeholder survey data collected from NCTR. In their role as partner and advocate for residency programs that follow a mentor model across the United States, NCTR sponsors annual surveys. Surveys go to residents, mentor teachers, principals in schools that use residents, and alumni of resident programs. Residency programs in NCTR's partner network administer the survey, and NCTR centralizes and reports on the data. We draw on survey data collected in 2016–17 and 2017–18 from residents and mentor teachers in 25 NCTR partner programs. (Once again, all teachers who host residents as a part of NCTR partner residency program follow a mentor model.)

To provide a description of co-teaching relationships and the types of joint activities in which they participate, we use data from 13 survey questions asked of residents and 15 questions asked of host teachers. Each question asks how often teachers and residents participate in specific activities with each other, using a five-point ordinal scale with the following categories: 1 (Never), 2 (one to two times per semester), 3 (monthly), 4 (weekly), and 5 (daily). Table 9 shows mean responses to the questions. Since mean responses to each of the questions did not vary much by year, we only present the mean response for data pooled across

both academic years. The data include 783 resident responses and 692 mentor teacher responses across 25 programs.

Residents and host teachers both reported fairly frequent participation in joint activities, with the average score being higher than three (i.e., monthly) on all questions. Both groups reported doing the following activities weekly or nearly weekly: examining strategies for classroom management, examining strategies for effective instruction, having dedicated meeting time, examining student progress, and examining how to adapt their teaching approach to meet students' learning needs. In general, teachers' responses were slightly higher than residents' responses. For example, host teachers reported more co-planning and co-teaching than residents reported, but both groups stated that these activities occur at least monthly, on average.

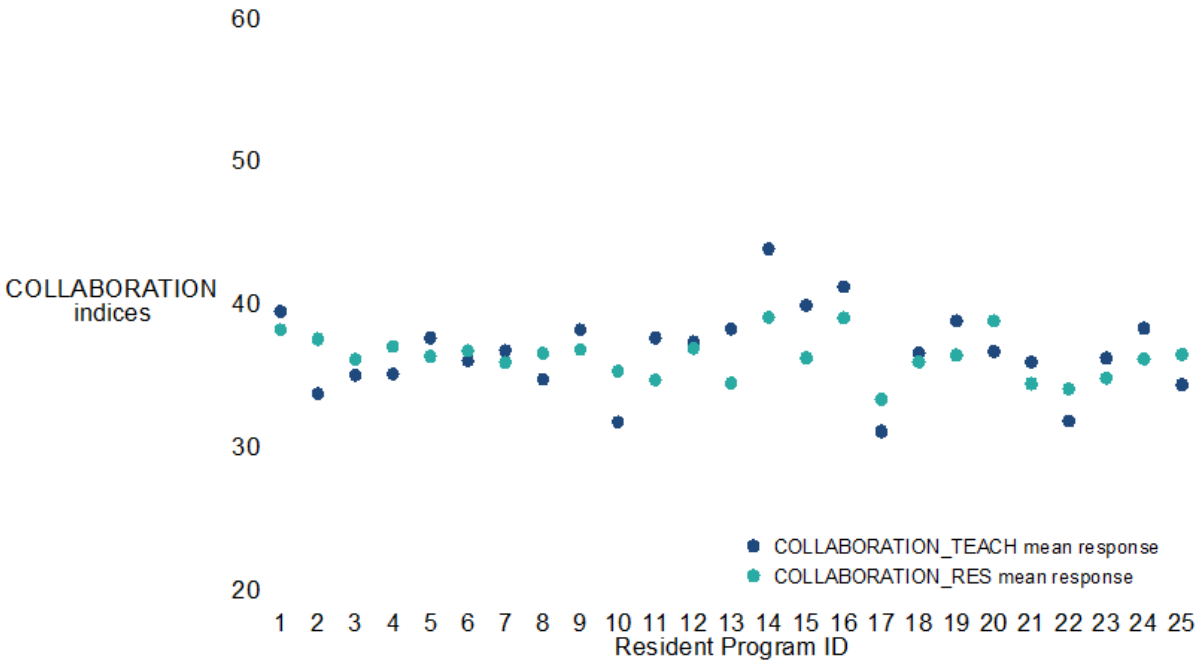
Though the overall averages are somewhat similar between teachers and residents, there was some variability across programs in the frequency with which host teachers and residents participate in these activities, such that teachers and residents in some programs reported participating in the activities more commonly than others. To look at this variation, we used principal components analysis to create a composite index of all 13 questions asked to residents. We call this index $\text{COLLABORATION}_{\text{RES}}$. We similarly computed a composite index variable to account for host teachers' responses to the 15 questions. We call this index $\text{COLLABORATION}_{\text{TEACH}}$. (See Appendix D for details on the principal components analysis methodology.) Figure 5 plots the scores for both groups. $\text{COLLABORATION}_{\text{RES}}$ varies across programs from 31 to 44 points, with an average of 37 points, demonstrating that there is minor variability across residency programs in the frequency with which residents report engaging in joint activities with their mentors. There was less variation in teachers' responses; $\text{COLLABORATION}_{\text{TEACH}}$ ranges from 33 to 39 points with an average of 36. Notably, for both the resident and teacher indices, program ID 14 has the highest index score while program ID 17 has the lowest score. As Figure 5 shows, teachers' and residents' average indices also seem to move together, suggesting they have fairly consistent views regarding the frequency with which they engage in joint activities.

Taken together, these results suggest that there are fairly high levels of coordination among teachers and residents in residency programs. Some programs had higher coordination across activities than others, indicating they may be somewhat better than others at encouraging collaboration between teachers and residents.

Table 9. Frequency with which residents and host teachers reported that they engaged in the following activities with each other, based on scale of 1 (never) to 5 (every day). Source: 2017 & 2018 NCTR Resident and Teacher survey.

Questions	Survey Years 2017 & 2018	
	Resident (Avg. score)	Mentor (Avg. score)
Examine strategies for classroom management	4.1	4.5
Examine strategies for effective instruction	4.0	4.3
Have dedicated meeting time	4.0	4.4
Examine the progress of students in your class	3.9	4.3
Examine how to adapt their teaching approach to meet student learning needs/styles	3.9	4.4
Co-teach	3.8	4.2
Examine how to assess student progress	3.7	4.1
Receive support to use new instructional approaches	3.7	4.2
Work to use multiple types of student data to inform planning and instruction	3.6	3.9
Co-plan instruction	3.6	4.1
Examine strategies to demonstrate professionalism and leadership	3.5	--
Use resident performance and effectiveness data to set instructional improvement goals for yourself	3.4	3.6
Examine strategies for effective student, family, and community engagement	3.3	3.6
Released full responsibility for all aspects of classroom instruction to your resident	--	3.7
Supported your resident to observe your practice	--	4.3
Supported your resident to promote diversity and inclusion in the classroom	--	4.3

Figure 5. Average COLLABORATION$_{RES}$ and COLLABORATION$_{TEACH}$ scores based on measures provided by teachers (in dark blue) and residents (in teal). Source: 2017 & 2018 NCTR Resident and Teacher survey.



To further understand how residents view the feedback, support, and mentorship they receive from their host teachers, we examine a set of questions asked to residents in the 2016–17 and 2017–18 surveys. The questions asked residents to indicate their agreement to statements regarding their host teachers' effectiveness and contribution towards their growth, with four ordinal response categories ranging from 1 (Strongly disagree) to 4 (Strongly agree). Table 10 shows mean responses to the questions across the two years of data. Mean responses to each of the questions did not vary much by year; therefore, we only present the mean response for data pooled across both academic years. The data include 783 resident responses across 25 programs. On average, residents tended to agree with all of the statements, with minor variation across questions. The residents' responses indicate that they generally feel supported and that they are receiving useful mentorship from their host teachers.

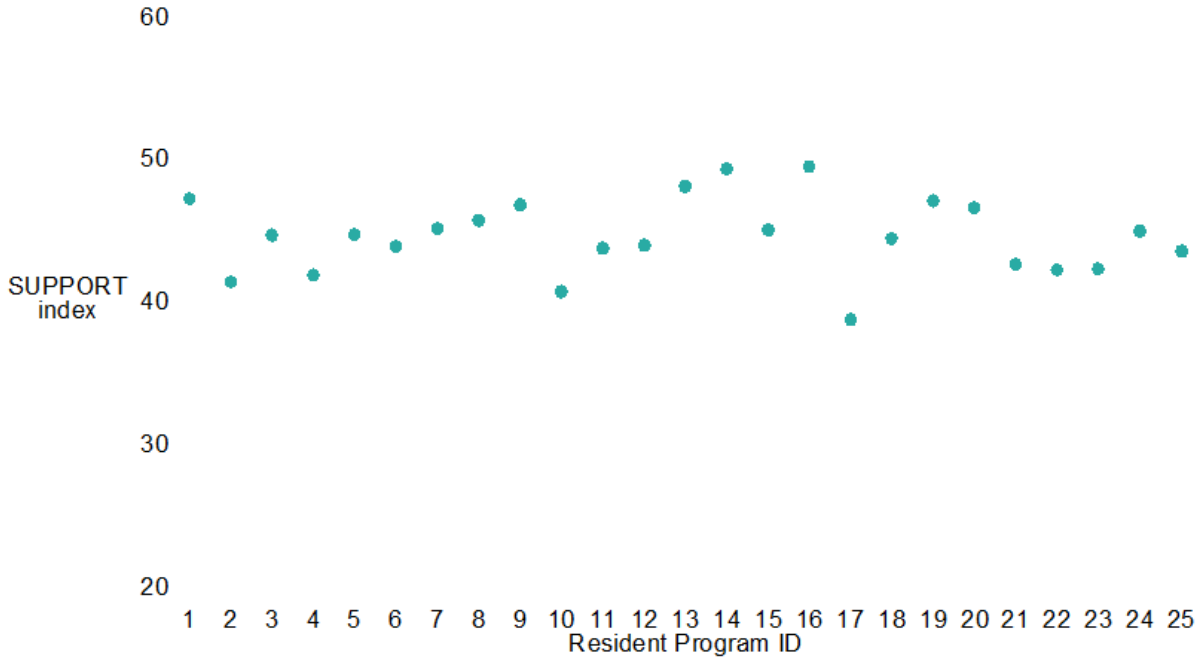
To examine variation in residents' responses across programs, we once again employed the principal components analysis method to construct a single composite index variable that would take into account residents' response to all 16 questions (see Appendix D for methodology). We call this new index variable SUPPORT. Figure 6 shows the SUPPORT index (averaged over both years) across all residency programs. There is some variation in the SUPPORT index scores across the residency programs; the index ranges from 39 to 49 with an average of 44. Notably, many of the programs that had higher scores on the COLLABORATION indices (e.g., program 14) also had higher scores on the SUPPORT index, and vice versa (e.g., see program 17).

Taken together, these data indicate that host teachers are providing anywhere from monthly to weekly coaching in a variety of areas to residents. On average, teachers and residents meet weekly, and teachers regularly share feedback with residents to improve their practice.

Table 10. Residents' agreement with the following statements regarding their host teachers' effectiveness and contribution towards their growth: "My current (or most recent) classroom mentor…" (Source: 2017 & 2018 NCTR Resident survey)

Questions	Survey Years 2017 & 2018 Avg. score
Is an effective teacher	3.6
Challenges me to grow	3.5
Gives me useful suggestions to improve my practice	3.4
Works with me to identify teaching challenges and possible solutions	3.4
Shares lesson plans, assessments, and other instructional activities	3.4
Gives me multiple opportunities to act on that feedback and improve my instruction	3.4
Makes me feel comfortable approaching my classroom mentor with questions and concerns	3.4
Encourages me to develop my individual teaching style	3.4
Identifies instructional goals and helps me develop realistic plans for achieving them.	3.3
Explains the rationale behind instructional decisions to me	3.3
Paces the gradual release of teaching responsibilities to me appropriately	3.3
Is a good match for me	3.3
Is an effective coach	3.3
Provides useful guidance on how to assess students informally on a daily basis	3.2
Gives me useful feedback on my lesson plans	3.2
Helps me apply what I am learning in my coursework	3.1

Figure 6. Average scores on the SUPPORT index based on residents' reported level of support from host teacher. Source: 2017 & 2018 NCTR Resident survey.



Are positive co-teaching relationships positively associated with student gains?

Assuming that strong relationships between host teachers and residents—as indicated by more co-planning and co-teaching—lead to better and more coordinated experiences in the classroom, we might expect that residents and teachers with strong relationships would be particularly beneficial to students. Specifically, we would expect that some of the positive relationship between having a resident in the classroom and improved student outcomes would be explained by these strong relationships. We hypothesize that residents and host teachers who report stronger relationships with each other will have students who perform better on assessments. To answer this question, we draw on NCTR stakeholder survey data from the two programs that partner with NCTR—Res Ed and City Teach—as well as student achievement data from their partner schools.

Data

We use NCTR resident survey from Res Ed and City Teach in 2017–18[9], as well as teacher surveys from 2016–17 and 2017–18 for both programs. We would have liked to merge student achievement data to the host teacher-resident pair data; however, the only identifiers in the survey data were residency program and school. We therefore conduct the analysis at the school level, aggregating the resident and host teacher composite indices (described above) and student achievement scores within schools. Using the school names, we merge the aggregated resident and host teacher data with the aggregated student achievement scores for each school in which host teachers and residents served. Table 11 below provides averages for the school

[9] We were only provided with school identifiers for the 2017–18 resident survey and hence we could only use the resident survey data from this year in our analysis.

proficiency rates and teacher and resident composite scores (represented by the $\text{COLLABORATION}_{\text{TEACH}}$ and $\text{COLLABORATION}_{\text{RES}}$ measures, respectively) for each sample included in the regressions below.

Table 11: School-level descriptive statistics. Source: Student-level academic records and NCTR teacher and resident surveys.

	All		Res Ed		City Teach	
Teacher Survey (2016–17 & 2017–18)	**Math**	**ELA**	**Math**	**ELA**	**Math**	**ELA**
$\text{COLLABORATION}_{\text{TEACH}}$	36.58	36.18	36.76	36.81	36.37	35.22
Math percent proficient (outcome year)	0.27	0.28	0.34	0.37	0.18	0.16
Math percent proficient (baseline)	0.28	0.30	0.38	0.40	0.17	0.16
ELA percent proficient (outcome year)	0.26	0.27	0.35	0.35	0.15	0.15
ELA percent proficient (baseline)	0.30	0.30	0.40	0.40	0.14	0.14
N	59	45	31	27	28	18
Resident Survey (2017–18)						
$\text{COLLABORATION}_{\text{RES}}$	34.55	34.55	35.67	35.67	33.81	33.81
Math percent proficient (outcome year)	0.28	0.28	0.42	0.42	0.18	0.18
Math percent proficient (baseline)	0.25	0.25	0.40	0.40	0.15	0.15
ELA percent proficient (outcome year)	0.30	0.30	0.49	0.49	0.17	0.17
ELA percent proficient (baseline)	0.28	0.28	0.48	0.48	0.15	0.15
N	25	25	10	10	15	15

Method

To examine the relationship between host teachers' composite scores and student growth, we fit the following multilevel regression model using two years of survey data:

$$\text{PrOutcome}_{jt} = \alpha_0 + \beta_0\, \text{Co-teaching}_{jt} + \phi_0\, \text{PrBaseline}_{jt} + \text{S}_j + \text{Y}_t + \varepsilon_{jt},$$

where PrOutcome_{jt} is the proficiency rate in math and English Language Arts (ELA) for school j in year t. Co-teaching_{jt} is the average composite index that captures teachers' perceptions of their co-teaching relationship with their resident, aggregated at the school level. PrBaseline_{jt} is the baseline proficiency rate in math and ELA respectively, for school j in year t, where baseline proficiency is the prior year proficiency in each subject. School-level characteristics might influence school proficiency rate and are likely to be correlated with response of teachers from the same school; therefore, we control for school fixed effects with S_j, which captures time-invariant, school-specific characteristics. We also control for year effects by including year fixed effects, captured by Y_t.

For the resident survey, we only have data for one year, so we conduct a simple ordinary least square regression to investigate if there is a correlation between school proficiency rates (in ELA and math) in the survey year and the composite index, controlling for school proficiency rates in the prior year.

Results

Table 12 below summarizes results from the multilevel regressions investigating the association between school proficiency rates and mentors' reported co-teaching relationship, followed by the results of the ordinary least square regressions investigating the association between school proficiency rates and residents'

reported co-teaching relationship. The first two columns present school math proficiency and ELA proficiency results for both programs, followed by results broken out by program (Res Ed in the second set of columns followed by City Teach in the last set of columns). There does not seem to be an association between the strength of the relationship between residents and teachers and proficiency in either subject, except for City Teach where we see a significant, positive association between residents' reported relationships and ELA proficiency rate. In particular, a one unit increase in the composite index for co-teaching relationship (using proxy 1) is associated with a one percentage point increase in the ELA proficiency rate, controlling for the baseline ELA proficiency rate. However, this is a small effect and, given that we see a significant association only in this case *and* that we would expect at least one of the effects to be significant by chance alone, we should not draw conclusions based on this result as it could just be driven by random noise in the data.

Table 12. Results from regression analyses examining relationship between aggregated co-teaching relationship variables and aggregated student proficiency. Source: Program records and NCTR mentor/resident survey data.

	All		Res Ed		City Teach	
	Math	ELA	Math	ELA	Math	ELA
Mentor Survey (2016–17 & 2017–18)						
Assoc. with $COLLABORATION_{TEACH}$	0.00	-0.01	0.00	-0.02	0.00	0.00
Significant?	No	No	No	No	No	No
N	59	45	31	27	28	18
Resident Survey (2017–18)						
Assoc. with $COLLABORATION_{RES}$	0.00	0.00	0.00	0.00	0.00	0.01
Significant?	No	No	No	No	No	Yes, p<.05
N	25	25	10	10	15	15
Assoc. with SUPPORT	0.00	0.00	0.00	0.00	0.00	0.00
Significant?	No	No	No	No	No	No
N	25	25	10	10	15	15

Do schools with a larger presence of residents show larger gains in student performance?

If students indeed benefit from having a resident in classroom, then schools with more classrooms with residents will potentially experience higher gains in student performance. We test this hypothesis below.

Data

To answer this research question, we use NCTR resident survey data from 2017–18 for Res Ed and City Teach. As above, we use school name to merge the resident survey to the school-level achievement data for each program. In the survey, residents are asked "How many residents teach in the same school as you?" with the option to choose from four categories: 0, 1, 2, and 3 or more. We add "1" to the value of each response to include the responding residents in the total. The new response categories thus range from 1 to 4 (since we do not know what number could be above 4, we cap the value of maximum response at 4). When there are multiple residents from a school answering this question in the survey, we confirm responses are consistent across residents. We then average the values reported by residents from the same school to create a measure of the total number of residents placed in each school. Table 13 provides averages for school proficiency rates and the number of residents placed in each school for each sample included in the regression analysis.

Table 13. School-level descriptive statistics showing number of residents in school and school proficiency rates. Source: Resident Survey (2017–18) and student-level academic records.

	All		Res Ed		City Teach	
	Math	ELA	Math	ELA	Math	ELA
No. residents in the school	2.40	2.40	1.90	1.90	2.73	2.73
Math prop. proficient (outcome year)	0.28	0.28	0.42	0.42	0.18	0.18
Math prop. proficient (baseline)	0.25	0.25	0.40	0.40	0.15	0.15
ELA prop. proficient (outcome year)	0.30	0.30	0.49	0.49	0.17	0.17
ELA prop. proficient (baseline)	0.28	0.28	0.48	0.48	0.15	0.15
N	25	25	10	10	15	15

Method

Using the one year of data, we once again conduct an ordinary least square regression to investigate if there is any correlation between school proficiency rates (in ELA and math) in the survey year and the number of residents in each school in the same year, controlling for school proficiency rates in the prior year.

Results

Table 14 summarizes the regression results. The first two columns present results for school math and ELA proficiency rates for both programs, followed by results broken out by program. There is no significant association between number of residents in a school and the school's proficiency rate in either subject in all cases, except for Res Ed, we see a significant negative association between number of residents in a school and the school's math proficiency rate. In particular, having an additional resident in a school is associated with a two percentage point decline in math proficiency rate, controlling for the baseline math proficiency rate, and the association is statistically significant. Although these results are contrary to our hypothesis that having more residents could lead to better academic outcomes for schools, it is difficult to draw any strong conclusions from these results given the very small sample size (N=10 for Res Ed) and the fact that we are drawing on only one year of data.

Table 14. Results from regression analyses examining relationship between number of residents in school and student proficiency. Source: Resident Survey (2017–18) and student-level academic records.

Resident Survey (2017–18)	All		Res Ed		City Teach	
Proficiency Outcome	Math	ELA	Math	ELA	Math	ELA
Association with no. residents in the school	-0.01	0.00	-0.02	0.00	-0.01	-0.01
Significant?	No	No	**	No	No	No
N	25	25	10	10	15	15

*** p<0.01, ** p<0.05, * p<0.1

LIMITATIONS

We note some limitations of the residency program analysis here. First, ideally we would have liked to collect and use data from school partners for the years preceding the introduction of the residency programs in the schools. This would have allowed us to establish a baseline trend for teacher effectiveness scores and further would have allowed us to match host teachers to non-host teachers based on this baseline trend. However, in some cases, districts were unwilling to share the data and, in other cases, schools began partnerships with residency programs in their first years, so there was no baseline. Second, TES is not measured in the same

way across programs; so as cautioned earlier, we cannot compare the impact across programs. Finally, to answer the supplemental research questions, we would have preferred to merge student achievement data to the host teacher-resident pair data so that we could conduct the analysis at a more granular level. However, at this point, in order to protect survey participants, the NCTR survey data does not include identifiers for residents or host teachers. We therefore conducted the analysis at the school-level, which meant we could not observe within-school heterogeneity. NCTR is working on creating datasets with identifiers that should prove useful for future research.

DIFFERENTIAL STAFFING PROGRAMS

DS programs are designed to provide alternative staffing models to schools with the primary goal of improving students' academic and/or behavioral outcomes. These programs, which are operated by independent nonprofit organizations, typically partner with schools that are under-resourced and/or serve economically disadvantaged students and aim to provide a service that schools cannot provide with their own staff. Some of the ways in which they do this include placing additional support staff (e.g., teaching assistants) in classrooms to lower the student-teacher ratio or by providing tutors who can offer personalized tutoring to students in school. DS programs typically recruit and train cohorts of recent college graduates to serve in this role. By design, these staff members usually serve finite terms—one or two years—after which they seek permanent employment elsewhere or attend graduate school.

Our hypothesis is that new teachers who previously served as differential staff members will be more effective than new teachers who did not have this experience. There are two reasons for this. First, DS programs operate as informal teacher pipeline programs by providing their staff members with intensive training in how to support students and also how to support a school's full-time staff (e.g., teachers, paraprofessionals, etc.). Differential staff members are able to work in a school setting, usually in a high-need school, and gain exposure to the challenges and rewards that accompany this experience. In a given week, staff members may work directly with students, take feedback and direction from teachers and administrators, plan and execute lessons, and experience the demands of working full-time in a school setting. These programs, therefore, provide hands-on training opportunities for young people who are interested in being educators. Some programs also forge partnerships with local graduate schools that allow staff members to simultaneously attend graduate school and receive their master's degree in education. In this way, would-be teachers gain valuable teaching experience before they ever become a full-time teacher.

Second, insofar that DS programs provide young people with hands-on experience in the classroom, they may also be effective at helping these young people self-identify whether teaching is the optimal career trajectory for them. While working in classrooms or schools, they may find they love the challenges they are facing and know that this is the right path for them to pursue. Alternatively, they may find that they do not enjoy teaching and that a different career either inside or outside of education would be preferable. Working as a differential staff member thus arms would-be teachers with additional information they can use as they make career choices. In this way, it promotes retention within the teaching profession of interested, committed teachers, while also helping weed out young people who are ultimately not interested in the profession. If we assume that interested, committed teachers are also more effective teachers, then this selection mechanism helps ensure that more effective teachers are making their way through the pipeline.

The primary research question addressed in this report is:

- Are teachers who worked with DS programs prior to becoming teachers more effective (as measured by student academic outcomes) than teachers who took more traditional routes to becoming teachers?

Additionally, we ask the following supplemental questions:

- What percentage of DS staff members go on to become teachers?
- Do DS alumni have higher retention rates than teachers who did not previously work for DS programs?
- Are cohorts of DS alumni teachers more diverse than teachers who did not previously work for DS programs?

To answer these questions, we draw on data from three DS programs: Teach Plus, Urban Corps, and Ed Serve. We provide background on each program below before explaining our data and methodology.

PROGRAM BACKGROUNDS

We began the DS portion of the study in a similar way to the residency portion. We entered an exploratory phase where we brainstormed potential program partners, conferred with experts (specifically those at Public Impact), and discussed with Overdeck relevant criteria on which we would want to select program partners (mainly programs that had already scaled or that had plans to scale). Overdeck began by suggesting two DS programs: Teach Plus and Urban Corps.

In our exploratory discussions with Teach Plus and Urban Corps, we identified the potential for smaller sample sizes than we would have preferred. Teach Plus operates in one large school district only (we call this district Fieldstone in this report), although they have grown rapidly and have a strong presence in the city. Urban Corps has a large presence in a second large district, where they launched their program, and a smaller presence in Fieldstone. We explored the feasibility of accessing data from Urban Corps' largest partner district in order to include staffing alumni from this district in our analysis. However, the district's data sharing requirements would have made it prohibitively difficult to access the necessary data. We proceeded with the intention to include Urban Corps-Fieldstone staffing alumni in the study but anticipated a small sample size since Urban Corps has a much smaller presence in Fieldstone. To bolster the number of programs included in the analysis, we proposed including Ed Serve as a third potential partner. Ed Serve is a national program; we include their Fieldstone program because we could efficiently request the data from the Fieldstone district's central office with Teach Plus and any alumni from Urban Corps, and we felt confident that their footprint in Fieldstone was large enough to yield a decent sample size for the study.

Below we briefly describe the three programs. Similar to the residency programs, to protect the anonymity of the organizations that participated in this study, we provide limited information on the background of the organization and their program model.

Teach Plus

Teach Plus is an education service model working in partnership with public high schools and middle schools in high-need neighborhoods in Fieldstone. Teach Plus provides math and ELA teachers with trained teaching assistants for an entire school year. Teaching assistants work alongside teachers within partner schools to lower the student-teacher ratio in classrooms as a means of providing students more individualized,

differentiated instruction. They also provide tutoring and extra support and facilitate extra learning time for students outside of the classroom. The majority of Teach Plus teaching assistants are recent college graduates (i.e., within three years of college graduation). They are trained by Teach Plus staff during a several week onboarding period and then receive ongoing PD throughout the year. After their first year of service, teaching assistants may decide to stay on for one additional year and may also decide to simultaneously enroll in a graduate education program with a local graduate school.

Ed Serve

Ed Serve is motivated by the assumption that schools have a surfeit of students who require extra academic and behavioral support services but are unable to offer these services because of resource constraints. Ed Serve aims to close this gap by placing extra staff members in high-need schools to provide supplemental academic and social-emotional support services. Ed Serve also provides after school programming at a subset of its partner schools. Ed Serve provides pre-service training, as well as ongoing PD to all staff members who are placed in schools. This includes a multi-week training experience leading up to their year of service. Staff members also have monthly PD, where they receive specialized training on a given topic related to teaching or general PD (e.g., resume writing, etc.). They can also attend ongoing PD throughout the year. Staff members may serve with Ed Serve for one or two years.

Urban Corps

Urban Corps was launched as a way of bringing high-dose tutoring to under-resourced schools. They focus predominantly on high schools not only because there is a need for services at this level but also because they see this time in a young person's educational career as a critical growth period. Urban Corps tutors provide one-on-one tutoring and coaching to students during a dedicated period each school day. Tutors typically stay with the program for one academic year (i.e., 10 months); although, some (often retirees) stay as long as three years. Urban Corps provides tutors with two weeks of training over the summer preceding the school year. During the school year, tutors receive ongoing training in the form of weekly observations, feedback from coaches, and weekly PD.

DATA AND STUDY SAMPLE

We answer our primary research question by studying student achievement among alumni of DS programs. We specifically study alumni who landed as teachers in Fieldstone schools, since both programs operate in Fieldstone, and Fieldstone offers a large sample size of teachers from which we could draw to create a comparison sample. To gather data from alumni, we first needed to know if they became teachers after leaving the respective DS program, as well as where and in what years they taught. We coordinated with each of the three DS program partners—Teach Plus, Urban Corps, and Ed Serve—to administer a survey for each program to collect data from their alumni. The primary purposes of the survey were to 1) determine who among the alumni went on to teach in non-charter Fieldstone schools[10] after they left the DS program, and 2) to obtain the necessary information (e.g., identifiers) to request teacher data from Fieldstone. We administered surveys for all three DS program partners. We used the survey responses of alumni who responded as having taught in a Fieldstone public school (after leaving their respective programs) to create a

[10] We limit to alumni teaching in non-charter public schools since Fieldstone does not provide data on charter school teachers.

roster of teachers to be submitted to Fieldstone. Fieldstone returned a data file that enabled us to identify most alumni from our roster, as well as their students. Table 15 summarizes the results of the survey and the number of alumni for whom we received data from the Fieldstone for each DS program partner.

The sample size for Urban Corps was ultimately too small to include in the analysis for the primary research question (though we do include them in some supplementary analyses). Thus, to answer the primary research question, we draw on data for two of the DS programs cited above: Teach Plus and Ed Serve.

Table 15: DS program survey results. Source: Alumni surveys.

DS Program	No. alumni surveyed	No. alumni responded	No. alumni responded they taught in Fieldstone	No. alumni for whom Fieldstone provided data
Teach Plus	106	43	27	27
Urban Corps	467	168	5	1
Ed Serve	2,012	365	32	19
Total	2,585	576	64	47

In our analysis, of the 47 alumni for whom we received data from Fieldstone, we include all who taught between 2014–15 and 2017–18 in subjects and grades that resulted in state assessment scores for their students. Table 16 provides a detailed breakdown of alumni teachers included in our study, broken out by high school (HS) versus elementary and middle school (EMS), as well as by the subject they taught. We include EMS teachers who teach math or ELA because students in these grades take the state math and ELA assessments. We also include HS teachers who teach math, ELA, or social studies because students in these grades take algebra exam, ELA, and history exams as a requirement for graduation. Below we discuss the study samples for Ed Serve and Teach Plus.

Ed Serve: As summarized in Table 16, our final sample of Ed Serve alumni teachers includes one Ed Serve alumni teacher who taught math to EMS students (giving us three observations since the same teacher taught in three different years); four Ed Serve alumni teachers who taught ELA to either EMS or HS students (giving us eight observations across years); and two Ed Serve alumni teachers who taught social studies to HS students (giving us seven observations across years). Thus, overall, we include seven unique Ed Serve alumni teachers who taught either math, ELA, or Social Studies in 2014–15 through 2017–18.

Teach Plus: Our final sample of Teach Plus alumni teachers includes seven Teach Plus alumni teachers who taught math to EMS or algebra to HS students (giving us 16 observations since the same teacher taught in multiple years); five Teach Plus alumni teachers who taught ELA to either EMS or HS students (giving us nine observations across years); and one Teach Plus alumni teacher who taught social studies to HS students (giving us three observations across years). Overall, we include 13 unique Teach Plus alumni teachers who taught either math, ELA, or social studies in academic years 2014–15 through 2017–18.

Table 16. Number of DS alumni by program, subject, and year. Source: Fieldstone records.

	Math teachers			ELA teachers			Social Studies teachers		
	Teach Plus (N)	Ed Serve (N)	All (N)	Teach Plus (N)	Ed Serve (N)	All (N)	Teach Plus (N)	Ed Serve (N)	All (N)
14–15 EMS	2	0	2	0	1	1	--	--	--
15–16 EMS	1	1	2	1	1	2	--	--	--
16–17 EMS	2	1	3	1	1	2	--	--	--
17–18 EMS	2	1	3	3	1	4	--	--	--
14–15 HS	1	0	1	1	0	1	1	2	3
15–16 HS	1	0	1	1	1	2	1	2	3
16–17 HS	3	0	3	1	1	2	1	2	3
17–18 HS	4	0	4	1	2	3	0	1	1
Total no.	16	3	19	9	8	17	3	7	10
Unique no.	7	1	8	5	4	9	1	2	3

Note: High School = HS, Elementary and Middle School = EMS

METHODOLOGY

We use a consistent methodology to conduct our analysis across both DS programs. We first match DS alumni teachers teaching EMS grades in math or ELA separately to non-DS alumni teachers from the Fieldstone schools who are also teaching EMS grades in each subject. Similarly, we match DS alumni teachers teaching HS grades in algebra, ELA and social studies separately to teachers from the Fieldstone schools who are also teaching HS grades in these subjects. We match each group of teachers on the following teacher- and school-level characteristics: years of teaching experience in Fieldstone public schools, average grade taught by teacher, and characteristics of the school at which they teach, including average math proficiency of school, average ELA proficiency of school, and percentage blacks or Latinx at school.

We use propensity score matching to match DS alumni teachers (treatment teachers) to similar teachers from the Fieldstone public schools (untreated teachers). We employ a nearest neighbor, one-to-many matching with replacement method, wherein each treatment teacher is matched to multiple untreated teachers that are closest in terms of propensity score. Since we are matching "with replacement," an untreated teacher can be used more than once as a match for each treatment teacher, which increases the average quality of matching.[11]

Table 17 below presents the characteristics of DS alumni teachers compared to teachers in the matched comparison group from the pooled sample (i.e., combining data from all four years for both HS and EMS grades) for each subject separately. The averages in the tables show that, for each subject, teachers in the DS group are very similar to those in the comparison group; there is no significant difference between the two groups on any of the characteristics as indicated by the fact that the means are very close and the p-values are greater than or equal to 0.70 in all cases.

Though the two groups of teachers are very similar on the matched characteristics, their students may still differ significantly on important factors such as prior test scores. Thus, once we achieve a good match for the teachers, we then match students taught by DS alumni teachers to students taught by comparison teachers using the same matching method we used for teachers (described above). Specifically, we match on the following characteristics: prior math and ELA state test scores, current class grade, IEP and ELL status of

[11] Caliendo, Marco, and Sabine Kopeinig, Some Practical Guidance for the Implementation of Propensity Score Matching. University of Cologne IZA Discussion Paper No. 1588. pp. 9, 2015. http://ftp.iza.org/dp1588.pdf

student, and whether they are black or Latinx. Table 18 below presents the characteristics of students taught by DS alumni, as well as the matched group of students taught by teachers in the matched comparison group. We show averages from the pooled sample (i.e., combining data from all four years for both HS and EMS grades) for each subject separately. As the table shows, for each subject, the matched student comparison groups are very similar to students of DS alumni; there is no significant difference between the two groups on any of the characteristics as indicated by the fact that the means are comparable and the p-values are high.

Our primary outcome measure is the state test score of students taught by teachers included in our sample, defined below (under "Dependent variable"). We run a multilevel regression model to examine whether the standardized state test scores are significantly different for students taught by DS alumni teachers compared to the matched student group, controlling for students' prior math and ELA state test scores, as well as whether the student has an IEP, ELL status, ethnicity, and class grade. We also control for the years of teaching experience of their teachers as well as school characteristics, including average school math and ELA proficiency and demographic characteristics. The model includes a binary variable indicating whether the teacher is a DS alumni teacher (1 if DS teacher, and 0 if not).

Table 17. Characteristics of DS teachers compared to teachers in the matched comparison group (pooled sample). Source: Fieldstone records.

	DS	Comp.	Diff.	p-value
Math				
No. years with Fieldstone	1.4	1.4	0.0	0.9
Avg. math proficiency of school (scale 1 to 4)	2.5	2.5	0.0	0.7
Pct. blacks or Latinx at school	89.2	89.7	-0.5	0.9
Avg. grade taught by teacher	8.2	8.1	0.1	0.8
N	19	171		
ELA				
No. years with Fieldstone	1.2	1.1	0.1	0.7
Avg. ELA proficiency of school (scale 1 to 4)	2.6	2.6	0.0	1.0
Pct. blacks or Latinx at school	82.0	82.4	-0.5	0.9
Pct. ELLs at school	18.3	17.8	0.5	0.9
Avg. grade taught by teacher	7.9	7.8	0.1	0.9
N	17	143		
Social Studies				
No. years with Fieldstone	2.3	2.4	0.0	0.9
Avg. math proficiency of school (scale 1 to 4)	3.0	3.0	0.1	0.9
Avg. ELA proficiency of school (scale 1 to 4)	3.0	3.0	0.0	0.9
Pct. blacks or Latinx at school	49.5	55.0	-5.4	0.7
Avg. grade taught by teacher	10.7	10.7	0.0	1.0
N	10	91		

Table 18. Characteristics of matched students taught by DS teachers compared to students taught by matched comparison teachers (pooled sample). Source: Fieldstone records.

	DS	Comp.	Diff.	p-value
Math				
Baseline math	-0.11	-0.13	0.03	0.55
Baseline ELA	-0.09	-0.11	0.02	0.68
Pct. IEP	0.25	0.27	-0.02	0.35
Pct. black	0.21	0.22	-0.01	0.45
Pct. Latinx	0.75	0.73	0.02	0.23
Current class grade	8.24	8.23	0.01	0.87
N	656	2843		
ELA				
Baseline math	-0.47	-0.46	-0.01	0.75
Baseline ELA	-0.35	-0.30	-0.05	0.17
Pct. IEP	0.23	0.23	0.00	0.90
Pct. ELL	0.15	0.12	0.03	0.38
Pct. black	0.28	0.28	-0.01	0.74
Pct. Latinx	0.62	0.61	0.01	0.76
Current class grade	8.83	8.83	0.00	0.98
N	897	3075		
History				
Baseline math	1.43	1.45	-0.02	0.64
Baseline ELA	1.18	1.15	0.03	0.51
Pct. IEP	0.06	0.07	-0.01	0.42
Pct. black	0.10	0.11	-0.01	0.71
Pct. Latinx	0.18	0.18	0.00	0.91
Current class grade	10.23	10.24	0.00	0.98
N	541	2145		

Dependent variable

The outcome variable is the student's standardized state test score in the outcome year (i.e., the year of analysis). We standardize the test scores within subject, year, and grade because in our analysis, we are including the following: 1) students in grades three to eight who take one set of assessments as well as students in high school who take different required exams; 2) student test scores from different years, and; 3) test scores from different subjects. Given the scale of test scores can vary across the type of exam, the year of exam, and the subject being tested, standardizing the scores places them on the same scale so that we can then interpret them in the same unit (i.e., standard deviation units).

Independent variables

Student level:

- **Prior year test scores**: Students' past performance on assessments is a major indicator of their future performance. In Fieldstone, third to eighth grade students are required to take ELA and math state assessments every year. In order to control for EMS students' prior performance, we therefore control for their standardized test score from their prior grade (e.g., using seventh grade ELA scores

for an eighth grade student). HS students, however, are not required to take a state assessment every year, and instead can choose to take their state exam once or multiple times while they are in HS. For all HS students, we therefore control for their eighth grade ELA and math state test score.

- **IEP status, ELL status, and ethnicity:** Students who have an IEP or are ELLs tend to score lower, on average, than students who are not in these groups. Likewise, there is a historical racial/ethnic gap in performance between blacks and Latinx and whites and Asians. We therefore account for any difference in the outcome variable due to these individual attributes by including them as controls in our models.
- **Current class grade:** We control for student grade in school, given that our sample includes students from all different class grades (fourth to 12th grades).

Teacher level:

- **DS teacher:** This is a binary variable capturing whether a teacher is a DS alumnus in a given year (1 if DS alumni, and 0 if not). The coefficient of this variable gives the mean difference between DS alumni teachers and comparison teachers on the outcome measure (i.e., the standardized test score). We similarly include binary variables for each DS program included in our analysis—Ed Serve and Teach Plus—to represent alumni affiliated with each program. The coefficient of each variable represents the mean difference between Ed Serve teachers versus their comparison teachers and Teach Plus teachers versus their comparison teachers.
- **Teaching experience (in years):** Teacher experience is captured by the total number of years a teacher has spent teaching in Fieldstone public schools. It is plausible that more experienced teachers are more effective teachers and that in itself could be associated with how students perform on assessments. It is, however, not clear whether teacher experience has a linear relationship with student performance, and we therefore introduce a quadratic term for years of experience.

School-level:

- A student's test score might be influenced by the characteristics of the school they attend. Therefore, we control for school-level characteristics such as average math and ELA proficiency across the school, percentage of students who are black or Latinx, percentage ELLs, and percentage of students with IEPs.

FINDINGS

For each subject, we pool together data from both EMS and HS grades from all years and run a regression on the pooled data to estimate the difference in performance between DS teachers' students versus comparison teachers' students. We run regressions separately for each subject. We also run a regression to separate out the effect for Teach Plus alumni teachers and Ed Serve alumni teachers. Finally, we combine data for all subjects together and run regressions to identify if there is an overall significant DS teacher effect.

Table 19 summarizes the main results of the analysis, first separately for each subject (first three columns), followed by the overall results pooled across all subjects (last column). (Appendix E includes detailed regression results for all subjects.)

Table 19. Results from regression analyses examining relationship between being a DS alumni teacher and student growth using dataset pooled across years. Source: Fieldstone academic records.

	Math (N = 3,499)	ELA (N = 3,971)	History (N = 2,686)	All Subjects (N = 10,156)
All DS alumni teachers	-0.15	-0.43	0.93	0.04
Comparison teachers	-0.13	-0.39	1.02	0.06
Diff	-0.02	-0.04	-0.09	-0.02
Sig?	No	No	No	No
Teach Plus alumni teachers	-0.15	-0.40	--	0.07
Comparison teachers	-0.13	-0.40	--	0.04
Diff	-0.02	0.00	--	0.03
Sig?	No	No	--	No
Ed Serve alumni teachers	--	-0.48	1.00	-0.03
Comparison teachers	--	-0.39	1.00	0.07
Diff	--	-0.09	0.00	-0.10
Sig?	--	No	No	No

*** p<0.01, ** p<0.05

Note: We do not include results for Ed Serve math since there is only one Ed Serve teacher in the sample. Likewise, we do not include results for Teach Plus history since there is only one Teach Plus teacher in the sample. However, we include all teachers (and students linked to the teachers) in the analysis for all subjects, so the results under "All Subjects" include both the one Ed Serve math teacher and the one Teach Plus history teacher.

Controlling for the variables in the model, the results for math (which are based on Teach Plus alumni teachers only) indicate that students of DS alumni teachers score slightly lower than students of comparison group teachers (0.02 standard deviation units), but the difference is statistically insignificant.

The results for ELA are similar. Students of DS alumni teachers score slightly lower than students of comparison group teachers (0.04 standard deviation units), but the difference is statistically insignificant. Among students of Ed Serve alumni teachers, the difference was -0.09 standard deviation units (statistically insignificant) between students of Ed Serve alumni teachers and students of comparison teachers; there was no difference between students of Teach Plus alumni teachers and students of comparison teachers.

The third column presents results for history (only Ed Serve alumni are reflected in this analysis). Students of DS alumni teachers also score slightly lower than students of comparison group teachers (0.09 standard deviation units), though again the difference is statistically insignificant.

In the last column, we present regression results on the pooled data for all subjects. Once again, we see no significant difference between the academic outcomes of students of DS alumni teachers compared to the outcomes of students of comparison group teachers. The results do not change when we break out the impact by DS program.

SUMMARY

The results do not support the hypothesis that teachers who previously worked with DS programs are more effective than similar teachers who did not work with a DS program. However, we have to be cautious about

drawing strong conclusions from these results because the sample size of unique DS alumni teachers is low (N=20 for all subjects; N=8 for math, N=9 for ELA, and N=3 for history).

SUPPLEMENTARY RESEARCH QUESTIONS

DS programs typically recruit recent college graduates and provide them with training and on-the-job experience, which could be in the form of offering additional support in classrooms alongside lead teachers or providing individualized tutoring to students, as a few examples. DS staff members typically serve for only one or two years in these programs. However, having received intensive training and exposure to the teaching profession from their program, many consider pursuing a teaching career. In this sense, DS programs may serve as informal teacher pipeline programs. Moreover, as explained earlier, to the extent that serving in a DS program exposes would-be teachers to the many challenges of teaching, this service may prompt some individuals to decide *not* to become teachers. We would expect that this dual mechanism—wherein interested, committed individuals are encouraged to pursue a career in teaching and less interested individuals are dissuaded from this decision—to have a net positive effect on teacher retention rates among DS alumni. We study below the percentage of DS alumni staff who go on to be teachers and then compare their retention rates to a comparison sample of teachers.

What proportion of DS front line staff members go on to become teachers?

We use data from the survey of DS program alumni of each DS program to estimate the percentage of DS alumni who become teachers. We asked respondents whether they taught in any school after they left the DS program. Table 20 shows the total number of alumni surveyed, the response rate (i.e., percentage of alumni who responded out of those surveyed), and percentage of alumni who became teachers out of those who responded. Overall, across all three DS programs, the response rate was at 48 percent, meaning 48 percent of alumni to whom a survey was sent responded to the survey. Among alumni who responded to the survey, 30 percent indicated they had become teachers. Broken out by programs, Teach Plus had the highest response rate (~95 percent), followed by Urban Corps (36 percent), and Ed Serve (~18 percent). Consistent with the response rate pattern, among those who responded to the survey, Teach Plus had the highest percentage of alumni who became teachers (~36 percent), followed by Urban Corps (~26 percent) and Ed Serve (~25 percent).

Table 20. Teacher pipeline rate among DS staff members who pursue a teaching career. Source: DS alumni survey.

	A	B	C	RESPONSE RATE	TEACHER PIPELINE RATE
DS Program	No. alumni surveyed	No. alumni responded	No. alumni responded they taught in any school	Pct. alumni responded out of those surveyed (B/A)	Pct. alumni who become teachers out of all who responded (C/ B)
Teach Plus*	290	274	98	94.5%	35.8%
Urban Corps	467	168	43	36.0%	25.6%
Ed Serve	2,012	365	91	18.1%	24.9%
All	2,769	807	232	47.8%	29.6%

*For Teach Plus, columns A to C are based on Teach Plus internal alumni survey they conducted in 2018. For this study, we only surveyed alumni who responded in the Teach Plus internal alumni survey that they taught in Fieldstone schools.

Do teachers who worked with DS programs prior to becoming teachers have higher retention rates than teachers who took more traditional routes to becoming teachers?

To answer this research question, we use three cohorts of teacher data (i.e., cohorts 2014–15, 2015–16, and 2016–17) from the three DS programs cited earlier: Teach Plus, Ed Serve, and Urban Corps. Each cohort is defined by the year in which they started teaching in a Fieldstone public school for the first time. We define a new cohort as teachers who are present in the teacher dataset for a particular year but missing in the teacher dataset in the previous year. As an example, to define teachers from the 2014–15 cohort, we only retain teachers who are in the 2014–15 teacher dataset but do not appear in the 2013–14 teacher dataset. For each cohort of teachers, we match DS alumni teachers to other teachers like them from the same cohort based on school-level characteristics and the length of time teaching in a Fieldstone public school using the matching method described previously (i.e., nearest neighbor, one-to-many matching with replacement).

In Table 21 we compare the average characteristics of the pooled sample of DS teachers (i.e., from all three cohorts) to that of comparison teachers. The averages in the table show that DS alumni teachers are very similar to those in the comparison group; there is no significant difference between the two groups in any of the characteristics as indicated both by the comparable means and the high p-values.

Table 21. Characteristics of DS alumni teachers compared to teachers in the matched comparison group (pooled sample). Source: Fieldstone records.

BASELINE VARIABLE	DS	Comp.	Diff.	p-value
No. years with Fieldstone	0.15	0.15	0.00	0.42
Avg. math proficiency of school (scale of 1 to 4)	2.30	2.34	-0.04	0.46
Avg. English proficiency of school (scale of 1 to 4)	2.40	2.43	-0.03	0.43
Pct. black or Latinx students at school	85.43	85.36	0.06	0.99
Pct. students with IEPs	22.04	21.39	0.65	0.73
Pct. students that are self-contained	2.47	2.55	-0.08	0.91
Pct. students that are ELLs	13.81	12.69	1.12	0.60
N	21	176		

Outcome

To answer the research question, we calculate teacher retention rate by tracking teachers from each cohort over time so we can calculate the number of years they have taught in Fieldstone public schools.[12] The latest year of data we have from the Fieldstone district is from 2017–18; we therefore use this year to calculate retention rate. As an example, for the 2014–15 cohort of teachers, retention rate is calculated as the number of years taught divided by the maximum number of years that a teacher could have taught (i.e., the dominator), which will be four since these teachers could only be tracked from 2014–15 (first year of service) through 2017–18 (the most recent year for which we have teacher data). For the 2015–16 cohort, the maximum number of years that a teacher could have taught is three, since these teachers could only be tracked from 2015–16 (first year of service) through 2017–18. Finally, for the 2016–17 cohort, the maximum number of years that a teacher could teach is two since these teachers could only be tracked from 2016–17 (first year of service) through 2017–18.

[12] One caveat is that if a teacher leaves the Fieldstone public school and goes on to teach in another district, we are not able to capture those teachers in our calculation of teacher retention rate.

Table 22 presents average retention rates for DS alumni teachers versus comparison group teachers for each cohort, as well as for the pooled sample. For each cohort, DS alumni teachers have a higher average retention rate compared to teachers in the comparison group, and the difference (10 and six percentage points for the 2014–15 and 2016–17 cohorts, respectively) is highly statistically significant. In the pooled sample, the retention rate of DS teachers is higher than that of the comparison group teachers by almost six percentage points, and the difference is highly statistically significant.

Table 22. Average retention rates for all DS alum teachers compared to comparison group teachers. Source: Fieldstone records.

	Retention rate				Ns	
Cohort year	**DS**	**Comp.**	**Diff.**	**Diff. sig.?**	**DS**	**Comp.**
2014–15	100.0%	90.0%	10.0%	***	4	40
2015–16	95.2%	92.9%	2.4%		7	46
2016–17	100.0%	94.0%	6.0%	***	10	90
All Cohorts	**98.4%**	**92.9%**	**5.6%**	***	**21**	**176**

We further break down the retention rates among DS teachers by program. Given that there is a very small representation of DS teachers in each cohort when we break the teachers by program, we do not conduct a significance test to examine whether the retention rate of DS teachers from each program is significantly different than that of the comparison group teachers. Table 23 shows that in the pooled sample, Ed Serve alumni teachers have a 100 percent retention rate and Teach Plus alumni teachers have a retention rate of 97 percent.

Table 23. Average retention rates for DS alum teachers from each DS program compared to comparison group teachers. Source: Fieldstone records.

	Retention rate for DS programs			Ns		
Cohort year	**Teach Plus**	**Ed Serve**	**Urban Corps**	**Teach Plus**	**Ed Serve**	**Urban Corps**
2014–15	100.0%	100.0%	--	2	2	--
2015–16	93.3%	100.0%	--	5	2	--
2016–17	100.0%	100.0%	100.0%	5	4	1
All Cohorts	97.2%	100.0%	100.0%	12	8	1

Is the population of DS alumni teachers more racially/ethnically diverse than the population of teachers who did not previously work for DS programs?

DS programs attempt to recruit ethnically diverse groups of young people to serve as teaching assistants or tutors in their programs. Consequently, we hypothesize that DS alumni who pursue a teaching career are more ethnically diverse than similar teachers who did not previously work for a DS program.

To answer this research question, we use teacher data from the three DS programs cited earlier: Teach Plus, Ed Serve, and Urban Corps. We conduct the analysis for the four most recent academic years: 2014–15 to 2017–18. For each academic year, we match DS teachers to other teachers from the district on school-level characteristics and their time in Fieldstone public schools, using the same matching method described earlier.

Table 24 compares the average characteristics of the pooled sample of DS teachers to that of matched comparison teachers. The averages in the table show that, in the pooled sample, teachers in the DS group are very similar to those in the comparison group; there is no significant difference between the two groups on any of the characteristics as indicated by the comparable means and high p-values.

We further break out DS teachers by program and conduct the matching for Teach Plus teachers and Ed Serve teachers separately. Tables 24 also compares the average characteristics of the pooled sample of Teach Plus alumni teachers to that of comparison teachers, followed by the average characteristics of the pooled sample of Ed Serve alumni teachers to that of comparison teachers. The averages in the tables confirm that in the pooled sample for each program, teachers in the DS group are very similar to those in the comparison group.

Table 24. Characteristics of DS teachers compared to teachers in the matched comparison group (pooled sample). Source: Fieldstone records.

	DS	Comp.	Diff.	p-value
ALL DS alum teachers				
No. years with Fieldstone	1.67	1.68	-0.01	0.96
Avg. math proficiency of school (scale of 1 to 4)	2.51	2.49	0.02	0.71
Avg. English proficiency of school (scale of 1 to 4)	2.63	2.61	0.02	0.64
Pct. black or Latinx students at school	80.73	81.87	-1.14	0.72
Pct. students with IEPs	21.13	21.02	0.11	0.91
Pct. students that are self-contained	2.35	2.34	0.01	0.97
Pct. students that are ELLs	13.48	14.64	-1.15	0.39
N	94	839		
Only Teach Plus alum teachers				
No. years with Fieldstone	1.53	1.53	0.00	0.99
Avg. math proficiency of school (scale of 1 to 4)	2.43	2.45	-0.01	0.79
Avg. English proficiency of school (scale of 1 to 4)	2.56	2.57	-0.01	0.75
Pct. black or Latinx students at school	86.58	86.34	0.24	0.94
Pct. students with IEPs	22.32	22.64	-0.32	0.72
Pct. students that are self-contained	2.59	2.67	-0.09	0.86
Pct. students that are ELLs	12.97	12.38	0.60	0.60
N	58	500		
Only Ed Serve alum teachers				
No. years with Fieldstone	1.98	2.02	-0.04	0.92
Avg. math proficiency of school (scale of 1 to 4)	2.66	2.66	0.00	0.98
Avg. English proficiency of school (scale of 1 to 4)	2.75	2.75	0.01	0.96
Pct. black or Latinx students at school	69.83	70.25	-0.43	0.95
Pct. students with IEPs	19.15	19.15	-0.01	1.00
Pct. students that are self-contained	2.04	1.81	0.22	0.67
Pct. students that are ELLs	15.02	15.90	-0.89	0.77
N	34	340		

*Two out of all DS teachers are Urban Corps teachers

Outcome

To answer the research question, we group the teachers into five primary racial/ethnic categories: black, Asian, white, Latinx, and other. We then compare the proportion of teachers that fall under each category between the DS alumni teachers and the comparison group teachers. The first block of Table 25 presents the comparison for all DS alum in the pooled sample, as well as for each individual year included in the analysis. In each academic year, with the exception of 2015–16, the proportion of DS alumni teachers who are Asian is significantly greater than the proportion Asian in the comparison group; likewise, the proportion of DS alumni teachers who are Latinx is significantly lower than the proportion Latinx in the comparison group. Consequently, in the pooled sample, DS alumni programs have a significantly higher proportion of Asian teachers (difference is 14 percentage points) but significantly lower proportion of Latinx teachers (difference is -10.5 percentage points). There is no significant difference between the two groups in the proportion of teachers who are white or black.

We further break out our analysis for Teach Plus and Ed Serve (the sample size for Urban Corps was too low to break out separately). Table 25 also provides a comparison of the proportion of teachers in each ethnic category, showing comparisons between Teach Plus alumni teachers (columns in the middle) and Ed Serve alumni teachers (columns in the right) and their respective comparison groups. These estimates reveal that the earlier results showing a significantly higher proportion of Asian teachers and a significantly lower proportion of Latinx teachers among all DS program teachers in comparison to comparison teachers are mainly driven by the ethnic composition of Teach Plus alum teachers. In the pooled sample, Teach Plus teachers have a significantly higher proportion of Asian teachers (difference is 20 percentage points) but significantly lower proportion of Latinx teachers (difference is -16 percentage points). Among Ed Serve teachers, the difference (roughly seven percentage points) in Asians is not statistically significant. Furthermore, there is virtually no difference in the proportion of Latinx teachers among Ed Serve teachers and the comparison group.

LIMITATIONS

There were some limitations of the DS analysis. First, the sample sizes were very low. This was expected given the nature of the study and the programs included in the study. The sample size for Urban Corps, for example, was too low to include the program in the analysis for the primary research question. Second, it is possible that the survey responses were biased if only certain types of DS alumni teachers responded to the survey (e.g., those who continue to serve as teachers and therefore want to participate in the study). This type of bias in the data could have implications for our findings on teacher retention rates. To improve future studies of this sort, programs should continue their alumni outreach efforts to ensure that they can track all alumni over time and promote inclusion in studies such as this one.

Table 25. Race/ethnic distribution for DS alum teachers compared to comparison group teachers. Source: Fieldstone records.

	All+					Teach Plus only					Ed Serve only				
	DS	Comp.	Diff.	p-value	Sig?	Teach Plus	Comp.	Diff.	p-value	Sig?	Ed Serve	Comp.	Diff.	p-value	Sig?
All Years (N)	94	839				58	500				34	340			
Blacks	14.9%	18.3%	-3.4%	0.39		12.1%	15.5%	-3.4%	0.46		20.6%	17.1%	3.5%	0.63	
Asians	21.3%	6.9%	14.4%	0.00	***	27.6%	7.2%	20.3%	0.00	***	11.8%	5.0%	6.8%	0.23	
Whites	50.0%	50.1%	-0.1%	0.98		48.3%	49.1%	-0.9%	0.90		50.0%	57.1%	-7.1%	0.43	
Latinx	11.7%	22.2%	-10.5%	0.00	***	8.6%	24.7%	-16.0%	0.00	***	17.6%	17.6%	0.0%	1.00	
Others	11.7%	14.6%	-2.9%	0.42		12.1%	13.8%	-1.7%	0.71		11.8%	16.8%	-5.0%	0.40	
2014–15 (N)	13	130				8	80				5	50			
Blacks	15.4%	16.2%	-0.8%	0.94		12.5%	22.5%	-10.0%	0.43		20.0%	10.0%	10.0%	0.60	
Asians	23.1%	3.8%	19.2%	0.11	*	37.5%	3.7%	33.8%	0.06	*	0.0%	2.0%	-2.0%	0.33	
Whites	53.8%	57.7%	-3.8%	0.79		37.5%	53.8%	-16.3%	0.37		80.0%	66.0%	14.0%	0.48	
Latinx	7.7%	19.2%	-11.5%	0.16		12.5%	17.5%	-5.0%	0.69		0.0%	18.0%	-18.0%	0.00	***
Others	0.0%	3.1%	-3.1%	0.05	**	0.0%	2.5%	-2.5%	0.16		0.0%	4.0%	-4.0%	0.16	
2015–16 (N)	19	170				12	100				7	70			
Blacks	21.1%	16.3%	4.7%	0.63		16.7%	14.2%	2.5%	0.83		28.6%	10.0%	18.6%	0.30	
Asians	21.1%	7.4%	13.7%	0.16		33.3%	5.0%	28.3%	0.05	**	0.0%	2.9%	-2.9%	0.16	
Whites	47.4%	60.0%	-12.6%	0.30		41.7%	50.0%	-8.3%	0.59		57.1%	65.7%	-8.6%	0.67	
Latinx	10.5%	13.7%	-3.2%	0.68		8.3%	27.5%	-19.2%	0.05	**	14.3%	12.9%	1.4%	0.92	
Others	47.4%	62.6%	-15.3%	0.21		41.7%	53.3%	-11.7%	0.45		57.1%	74.3%	-17.1%	0.39	
2016–17 (N)	28	242				16	130				11	110			
Blacks	14.3%	16.4%	-2.1%	0.76		18.8%	15.0%	3.8%	0.72		9.1%	17.3%	-8.2%	0.39	
Asians	21.4%	5.4%	16.1%	0.05	**	25.0%	10.0%	15.0%	0.19		18.2%	3.6%	14.5%	0.22	
Whites	50.0%	50.4%	-0.4%	0.97		50.0%	48.1%	1.9%	0.89		45.5%	56.4%	-10.9%	0.49	
Latinx	14.3%	27.9%	-13.6%	0.06	*	6.2%	26.9%	-20.6%	0.01	***	27.3%	22.7%	4.5%	0.75	
Others	--	--	--	--		--	--	--	--		--	--	--	--	
2017–18 (N)	34	297				22	190				11	110			
Blacks	11.8%	21.2%	-9.4%	0.12		4.5%	14.1%	-9.5%	0.07	*	27.3%	24.5%	2.7%	0.85	
Asians	20.6%	9.4%	11.2%	0.13		22.7%	7.7%	15.0%	0.11		18.2%	9.1%	9.1%	0.45	
Whites	50.0%	41.8%	8.2%	0.37		54.5%	47.7%	6.8%	0.55		36.4%	48.2%	-11.8%	0.44	
Latinx	11.8%	23.5%	-11.8%	0.06	*	9.1%	24.1%	-15.0%	0.03	**	18.2%	15.5%	2.7%	0.82	
Others	5.9%	4.1%	1.8%	0.67		9.1%	6.4%	2.7%	0.67		0.0%	2.7%	-2.7%	0.08	*

*** p<0.01, ** p<0.05, * p<0.1

+Two out of the total DS teachers are Urban Corps teachers

SUMMARY OF KEY FINDINGS

This report provides evidence that, in some settings, there are ancillary benefits of residency programs and DS programs. Below we present a summary and short discussion of key findings for both programs.

Residency Programs

We explore whether hosting a resident in the classroom improves teacher effectiveness, as measured by TES. We found a significant, positive effect for two residency programs (i.e., Res Ed and City Teach), where hosting a resident in the classroom was associated with improvements in TESes in the full sample (i.e. all years pooled together). We did not find a significant effect of hosting a resident in the classroom for Teach Plus. This finding is interesting, and in many ways logical, because Teach Plus uses a different program model than Res Ed and City Teach. As described earlier, Res Ed and City Teach use the mentor model, wherein experienced, high performing teachers are selected to host residents in the classroom. These programs also provide significant coaching and support to the mentors as they, in turn, learn to work with and support the residents. As the analyses in Tables 9 and 10 show, in the mentor model, both mentor teachers and residents report fairly high levels of collaboration: on average, they meet at least weekly; they regularly discuss strategies for effective instruction and classroom management; they co-plan, and; they draw on student data to inform instruction. All of these actions could help drive student learning.

Teach Plus, on the other hand, uses the host model. They select host teachers who are willing to accommodate a resident. They do not train or coach the host teachers and the host teachers are not expected to provide significant coaching to the residents. We see this difference in teacher recruitment reflected in the descriptive data provided in Table 7. Teach Plus host teachers have roughly the same experience and prior effectiveness scores as non-host Teach Plus teachers, whereas host teachers at Res Ed and City Teach are generally more experienced and have higher prior effectiveness scores than the comparison group.

It is possible that differences in the program models—the ways these programs recruit, train, and deploy teachers—can explain some of the differences in the observed effects on student learning. If the mentor model leads to closer coordination and collaboration between the host teacher and resident within the classroom, then this could lead to improved student outcomes. We attempted to explore this with one of our supplemental analyses; specifically, we examined whether teachers and residents with stronger reported co-teaching relationships also had more growth in teacher effectiveness. We did not find a consistent relationship between our measure of strong relationships (taken from the stakeholder surveys) and teacher effectiveness. Since we had to perform this analysis at the school level, it is unsurprising that we did not observe a strong effect. We did find one instance of a significant, positive (albeit, modest) association between City Teach residents' perceptions of the co-teaching relationship and school ELA proficiency rate. A one unit increase in the composite index for co-teaching relationship was associated with a one percentage point increase in the ELA proficiency rate, controlling for the baseline ELA proficiency rate. Nonetheless, additional work is needed to unpack why this mentor model might lead to more effective teaching.

Two additional considerations are important to highlight. First, we found no evidence that hosting a resident in a classroom was associated with lower TESes for host teachers. This is important, given the demands of accommodating, let alone mentoring, another adult in the classroom. It is plausible for teachers to be *less* effective when a resident was in the classroom (through the mechanisms we discuss earlier). However, we found no evidence of this. Our results indicate that teachers are either as effective (in the case of Teach Plus) or more effective (as in the case of City Teach and Res Ed) when hosting a resident in the classroom.

Moreover, these results say nothing about the programs' effectiveness at meeting their primary goals of training future teachers via the residency model. We find evidence that Teach Plus teachers who hosted residents were as effective as teachers who did not host residents, but this says nothing about the program's effectiveness at training future teachers. Likewise, we found some evidence that Res Ed and City Teach teachers are more effective when they host residents in the classroom, but separate analyses are required to test their effectiveness at training future teachers.

Differential Staffing Programs

We were primarily interested in learning more about the extent to which DS programs create cohorts of effective teachers. We did not find evidence that alumni of DS programs are more effective teachers than teachers who did not serve with the DS programs. Student test scores did not differ significantly between DS alumni teachers' students and students of a carefully matched comparison group of teachers. However, we did find that DS alumni are more likely to persist at teaching: retention rates were roughly six percentage points higher for the DS alumni than they were for the matched comparison group.

We also find that, out of all alumni who responded to the survey, approximately 36 percent of Teach Plus alumni, 26 percent of Urban Corps alumni, and 25 percent of Ed Serve alumni went on to become teachers. With respect to the accuracy of our results, Ed Serve had previously surveyed their alumni and learned that approximately 20 percent pursued a career in teaching, which was close to our estimate of 25 percent. Moreover, the estimates for Teach Plus are comparable to what they reported to us based on their anecdotal knowledge about their alumni. The fact that the Urban Corps estimates are in the same range as the other program estimates gives us some confidence that they are reliable.

Regarding the question of whether the DS alumni are more diverse than non-DS alumni, our results suggest that DS alumni who go on to be teachers are more likely to be Asian and less likely to be Latinx. These results are mainly driven by the ethnic composition of Teach Plus alumni teachers: there seems to be no significant difference in the ethnic composition of Ed Serve teachers and comparison teachers.

APPENDIX A: Relate teacher evaluation methodology.

Relate's teacher evaluation program takes a multi-measure approach for instructional staff where two of the goals are effort based and two goals are outcome based. In the first table below, Goals 1 and 2 are outcome-based goals tied to assessments while the full observation and EOY evaluation goals are effort based. The first table illustrates the model for our K–12 teachers. The second table illustrates how the points earned for each goal are then used to determine overall performance calculations. (This information was obtained directly from Relate.)

Goals	Possible Points	Points Received
Goal #1 MAP Reading: % of scholars meeting their growth goal	1 Point – Fewer than 50% 2 Points – 50% to 69% 3 Points – 70% to 79% 4 Points – 80% or greater	2
Goal #2 MAP Math: % of scholars meeting their growth goal	1 Point – Fewer than 50% 2 Points – 50% to 69% 3 Points – 70% to 79% 4 Points – 80% or greater	3
Goal #3: Classroom Observation	1 Point – Below Expectations 2 Points – Meets Expectations 3 Points – Exceeds Expectations 4 Points – Exemplar	2
Goal #4: End of Year Evaluation	1 Point – Below Expectations 2 Points – Meets Expectations 3 Points – Exceeds Expectations 4 Points – Exemplar	2
	Total Points Received	9

With 9 points this employee is "Meeting Our High Bar" and receives a raise		
7 or fewer = Working Towards our High Bar	Employee in a developmental stage and can expect additional support in order to meet or exceed our high bar.	X% Raise
8–12 = Meeting our High Bar	Team members are proficient across most/all evaluation categories, and achieving solid results.	Y% Raise
More than 13 points	Team members at this level set the tone for what excellence looks like.	Z% Raise

APPENDIX B: Characteristics of host teachers compared to teachers who did not host (by academic year).

Source: Program academic datasets.

Res Ed

	2012–13		2013–14		2014–15		2015–16		2016–17	
	Host	Other	Host	Other	Host	Other	Host	Other	Host	Other
Teacher's TES in Baseline year/s	3.41	2.92	3.49	3.20	3.64	3.25	3.69	3.28	3.67	3.34
Years teaching experience	5.75	4.67	6.47	5.20	6.15	5.27	6.40	5.22	7.54	5.64
Teacher is a female	84.62	82.67	80.77	77.78	86.67	77.34	84.62	75.18	84.91	75.67
School level characteristics										
Pct. students with IEPs	1.31	1.05	1.26	0.98	0.73	1.00	1.18	1.25	1.82	1.78
Pct. LEP students	29.55	24.77	31.81	28.98	25.62	27.03	18.71	21.00	17.08	20.01
Pct. Free and reduced-price lunch students	80.03	76.19	82.82	74.79	79.79	79.04	83.60	81.29	76.88	78.45
Pct. Latinx and black students	86.08	81.06	88.75	82.17	91.29	85.09	88.96	86.47	85.27	85.65
Pct. white and Asian students	10.30	15.05	8.39	13.81	6.00	11.19	8.18	9.96	11.12	10.15
No. teachers	26	202	26	315	30	331	52	411	53	448

City Teach

	2013–14		2014–15		2015–16		2016–17		2017–18	
	Host	Other	Host	Other	Host	Other	Host	Other	Host	Other
Teacher's TES in Baseline year/s	1.95	0.95	0.68	0.46	0.22	-0.08	1.06	0.24	1.04	0.07
School level characteristics										
Pct. students with IEPs	9.61	11.79	10.77	11.78	11.71	10.85	12.73	11.01	14.14	10.94
Pct. Economically disadvantaged students	87.99	80.06	87.60	78.70	90.62	78.89	93.00	81.48	86.47	77.70
Pct. Latinx and black students	92.23	86.88	90.44	87.38	92.47	87.28	91.43	89.45	92.90	87.52
Pct. white and Asian students	7.10	11.86	8.69	11.39	6.18	10.53	10.25	12.45	5.10	9.76
No. teachers	18	1820	29	1723	13	700	26	1487	27	1560

Teacher Prep

	2016–17		2017–18	
	Host	Other	Host	Other
Teacher's TES in Baseline year/s	2.39	2.26	2.04	2.24
Years teaching experience	3.75	4.22	3.04	4.71
Teacher's ethnicity				
Black	27.08	28.36	28.30	25.77
Asian	2.08	5.18	7.55	5.93
Latinx	4.17	3.11	1.89	2.45
Other	4.17	2.07	1.89	1.84
White	62.50	61.28	60.38	64.01
School level characteristics				
Pct. students with IEPs	6.82	6.52	7.42	6.71
Pct. LEP students	36.68	30.99	34.15	31.84
Pct. Free and reduced-price lunch students	82.00	71.50	83.83	73.03
Pct. Latinx and black students	93.26	82.51	94.95	82.88
Pct. white and Asian students	5.20	15.45	3.63	14.92
No. teachers	48	483	53	489

APPENDIX C: Residency program regression results from model specification 1. Source: Program records.

Table C1. City Teach

	All years	2013–14 only	2014–15 only	2015–16 only	2016–17 only	2017–18 only
City Teach host	0.60***	0.78	0.72**	0.41	0.54*	0.17
	(0.182)	(0.545)	(0.325)	(0.345)	(0.310)	(0.238)
TES baseline	0.33***	0.41***	0.44***	0.14***	0.27***	0.21***
	(0.011)	(0.020)	(0.020)	(0.038)	(0.022)	(0.017)
School level characteristics						
% SPED	-0.02***	-0.02	-0.01	-0.05***	-0.02	0.01
	(0.007)	(0.022)	(0.012)	(0.017)	(0.014)	(0.010)
% Free and reduced-price lunch	-0.00	-0.02*	0.01**	0.00	0.00	-0.01*
	(0.003)	(0.009)	(0.004)	(0.008)	(0.006)	(0.004)
% Latinx and black	0.03*	0.01	0.13*	-0.02	-0.16**	0.02
	(0.016)	(0.103)	(0.081)	(0.066)	(0.067)	(0.037)
% white and Asian	0.03*	-0.01	0.14*	-0.01	-0.13**	0.02
	(0.017)	(0.109)	(0.086)	(0.072)	(0.062)	(0.040)
year2	0.32***					
	(0.053)					
year3	-0.05					
	(0.076)					
year4	-0.38***					
	(0.078)					
year5	0.08					
	(0.064)					
Constant	-2.74*	0.22	-13.79*	1.74	15.49**	-1.86
	(1.626)	(10.307)	(8.102)	(6.644)	(6.760)	(3.603)
Observations	7,403	1,838	1,752	713	1,513	1,587
Number of groups	178	146	139	117	139	139

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table C2. Res Ed

	All years	2012–13 only	2013–14 only	2014–15 only	2015–16 only	2016–17 only
Res Ed host	0.11***	0.21***	0.08	0.11	0.07	0.06
	(0.037)	(0.074)	(0.080)	(0.085)	(0.069)	(0.064)
TES baseline	0.50***	0.47***	0.59***	0.55***	0.57***	0.53***
	(0.024)	(0.045)	(0.049)	(0.052)	(0.046)	(0.044)
Yrs. Teaching experience	0.02**	-0.03	-0.00	-0.01	0.03**	0.03***
	(0.007)	(0.019)	(0.015)	(0.016)	(0.012)	(0.011)
Yrs. Teaching experience * Yrs. Teaching experience	-0.00**	0.00	-0.00	0.00	-0.00**	-0.00***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)
School level characteristics						
% SPED	-0.00	-0.01	-0.01	0.06*	0.06**	-0.03
	(0.009)	(0.027)	(0.023)	(0.031)	(0.025)	(0.019)
% Free and reduced-price lunch	-0.00	-0.01**	0.00	0.00	0.00	0.01
	(0.002)	(0.005)	(0.004)	(0.002)	(0.008)	(0.006)
% Latinx and black	0.02**	0.00	0.04***	0.03**	0.02	0.02
	(0.007)	(0.012)	(0.010)	(0.013)	(0.016)	(0.014)
% white and Asian	0.02**	-0.01	0.05***	0.04**	0.03*	0.03*
	(0.008)	(0.014)	(0.012)	(0.016)	(0.018)	(0.018)
year2	-0.09**					
	(0.036)					
year3	0.07*					
	(0.037)					
year4	0.03					
	(0.037)					
year5	0.13***					
	(0.038)					
Constant	0.35	2.90**	-2.16**	-1.14	-0.69	-0.77
	(0.668)	(1.134)	(0.980)	(1.243)	(1.373)	(1.393)
Observations	1,894	228	341	361	463	501
Number of groups	42	33	34	37	39	39

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table C3. Teacher Prep

VARIABLES	All years	2016–17 Only	2017–18 Only	All years
Teach Plus host	-0.02	-0.04	-0.01	--
	(0.055)	(0.073)	(0.078)	
No. months worked as Teach Plus host				-0.01
				(0.010)
TES baseline	0.30***	0.29***	0.41***	0.31***
	(0.031)	(0.035)	(0.050)	(0.031)
Teaching experience	0.01	-0.01	0.02*	0.01
	(0.008)	(0.010)	(0.010)	(0.008)
Teaching experience * Teaching experience	-0.00	0.00	-0.00	-0.00
	(0.000)	(0.000)	(0.000)	(0.000)
School level characteristics				
% SPED	0.01	-0.02	0.00	0.01
	(0.014)	(0.022)	(0.021)	(0.014)
% Free and reduced-price lunch	0.01	0.01	0.01	0.01
	(0.007)	(0.010)	(0.011)	(0.007)
% Latinx and black	0.03	0.02	0.02	0.02
	(0.024)	(0.046)	(0.030)	(0.024)
% white and Asian	0.04	0.04	0.04	0.04
	(0.024)	(0.045)	(0.030)	(0.024)
year2	-0.19***			-0.19***
	(0.031)			(0.031)
Constant	-2.00	-1.51	-2.50	-1.82
	(2.277)	(4.247)	(2.813)	(2.270)
Observations	1,073	531	542	1,060
Number of groups	36	33	36	36

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

APPENDIX D: Principal components analysis

Given the residents are asked 13 different questions, all of which essentially try to capture the extent to which residents and host teachers are engaging in joint activities, we employed the principal components analysis method to construct a single composite index variable using the information from the 13 questions (variables). The method allows us to estimate the optimal weight to assign to each question, such that we are capturing as much information from the original questions as possible based on the correlations among those variables. Since all the 13 variables were measured in the same frequency scales, we simply applied the weights derived from the factor loadings associated with the first factor (which explained 57 percent of the variance) and applied the weights to each question to compute the composite index. The index variable can, therefore, be thought of as a weighted average of the 13 original questions or variables, where weights are the factor loadings. Using the composite index is an efficient way to capture whether residents' answers to the different questions reflected that they were more frequently working together with their mentor.

As with residents, in order to capture what host teachers say about engaging in joint activities with their residents (which, again, can be thought of as a proxy for the quality of the co-teaching relationship) we focus on a set of 15 questions that the host teachers were asked in the teacher survey in academic years 2016–17 and 2017–18. As with the questions asked to residents, we once again employed the principal components analysis method to construct a single composite index variable that would take into account teachers' response to all the 15 questions (variables). Since all 15 variables were measured in the same frequency scales, we simply applied the weights derived from the factor loadings associated with the first factor (which explained 33 percent of the variance) and applied the weights to each variable to compute the composite index. The index variable can, therefore, be thought of as a weighted average of the 15 original questions, where weights are the factor loadings.

Lastly, we employed the principal components analysis method to construct a single composite index variable that would take into account residents' responses to all the 16 questions (variables) gauging their agreement with statements regarding their mentor's effectiveness and contribution towards their growth. Since all 16 variables were measured in the same frequency scales, we simply applied the weights derived from the factor loadings associated with the first factor (which explained 69 percent of the variance) and applied the weights to each variable to compute the composite index. The index variable can, therefore, be thought of as a weighted average of the 16 original questions, where weights are the factor loadings.

APPENDIX E: DS program regression results for all subjects. Source: Fieldstone records.

Variables	All Subjects for all DS programs	All Subjects for Ed Serve vs. Teach Plus
DS teacher	-0.02	--
	(0.041)	--
Ed Serve		-0.1
		(0.055)
Teach Plus		0.03
		(0.050)
Prior math	0.25***	0.25***
	(0.020)	(0.020)
Prior ELA	0.29***	0.29***
	(0.027)	(0.027)
Other student-level characteristics		
IEP	-0.27***	-0.27***
	(0.034)	(0.034)
ELL	-0.38***	-0.38***
	(0.061)	(0.061)
Black	-0.18***	-0.18***
	(0.029)	(0.029)
Latinx	-0.16***	-0.16***
	(0.033)	(0.032)
Other	0.01	0.01
	(0.052)	(0.052)
White	-0.02	-0.02
	(0.023)	(0.023)
Grade in current year	0.04**	0.04**
	(0.021)	(0.021)
School and teacher-level characteristics		
Total years with district	0.03	0.03
	(0.021)	(0.022)
Total years with district * Total years with district	-0.00	-0.00*
	(0.002)	(0.002)
Average math proficiency	0.40**	0.39**
	(0.193)	(0.190)
Average English proficiency	0.03	0.03
	(0.246)	(0.244)
% black or Latinx in school	0.00	0.00
	(0.001)	(0.001)
% ELL in school	0.01***	0.01***
	(0.002)	(0.002)
% students with IEPs in school	0.00	0.00
	(0.003)	(0.003)
Constant	-1.21***	-1.21***
	(0.351)	(0.350)
Observations	10,156	10,156
Number of groups	129	129

Robust standard errors in parentheses
*** p<0.01, ** p<0.05