Estimating an Ancillary Benefit of Teacher Residency Programs

Rebecca Casciano*

Pratikshya Bohra-Mishra

Jennifer Puma

Glass Frog Solutions

*Direct correspondence to Rebecca Casciano at rebecca@glassfrog.us. 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. 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. Senad Lekpek provided considerable research assistance on this project. John Friedman provided valuable technical consultation. All mistakes and omissions belong to the authors.

Abstract

This paper explores the ancillary benefit of teacher residency programs, which 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 a resident in the classroom, there are positive, short-run benefits to student learning for the students in the classroom. We use data from three teacher residency programs to explore whether hosting a resident in the classroom improves teacher effectiveness scores among host teachers, 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 teacher effectiveness scores being higher for host teachers than comparison teachers at two of the programs. The two programs for which we detect a significant, positive effect employ a resident training approach called the "mentor model," in which host teachers receive training on how to mentor the residents in the classroom, including but not limited to how to co-plan, collaborate, and co-teach with the residents. The program in which we do not find a significant effect used a different approach in which host teachers simply host the resident but are not expected to mentor or train them throughout the year. We discuss the implications of these different approaches for teacher effectiveness and offer and test some hypotheses for why the mentor model may lead to improve teacher effectiveness during the years a teacher is hosting a resident.

INTRODUCTION

Teacher residency programs aim to recruit and train teachers to serve in under-resourced school districts as a means of addressing the nationwide shortage of talented teachers in these districts. Residency programs combine theoretical training with experiential learning, immersing teachers-in-training (i.e., residents) in classrooms to learn how to teach and manage a classroom alongside an experienced teacher in a live setting. Residency programs often place graduating residents in schools in the same district upon obtaining their teaching certification. The underlying hypothesis is that the residency experience provides more realistic, organic opportunities for teachers-in-training to learn and practice teaching and that this extra preparation will translate into greater effectiveness and higher retention rates in the long-term.

Though their primary goal, in the long run, is to recruit and train high quality teachers, residency programs may also have shorter-term, ancillary benefits on teacher effectiveness and, hence, student learning. The present paper draws on data from three residency programs to test whether hosting a resident is associated with significantly greater teacher effectiveness in the year(s) in which teachers host a resident. Below we describe residency programs in greater detail, the mechanisms by which placing a resident in a classroom could ultimately improve student learning, and why these ancillary benefits may be relevant for districts hosting residents.

Residency programs

Residency programs offer an innovative approach to meet the teacher recruitment, training, and retention challenges faced by under-resourced school districts, as well as hard-to-staff geographic areas and academic subjects such as mathematics, science, special education,

and English as a second language (Guha, Hyler, & Darling-Hammond, 2017; Solomon, 2009). 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. Residency programs build on research suggesting that new teachers who are placed in schools that are demographically similar to the schools in which they were trained are ultimately more effective than new teachers placed in dissimilar schools (Goldharber, Krieg & Theobald, 2017). They thus offer a unique pipeline by which under-resourced school districts can recruit and retain highly effective teachers who will be ready to serve students as soon as they earn their teaching certificates. In this way, residency programs may be well-positioned to meet the human capital needs of school districts in ways that traditional teacher training programs or schools of education have not been able to (National Center for Teacher Residencies, 2015; Solomon, 2009; Rosenberg and Miles, 2018).

Modeled after medical residencies, teacher residency programs provide pre-professional, experiential learning opportunities to residents; residents typically participate in a year-long apprenticeship where they are paired with a classroom teacher and work in that teacher's classroom during the full school year (Solomon, 2009). In many residency programs, residents typically begin the year observing, co-teaching, and in some programs, being coached by the host teacher; as the year goes on, the residents gradually lead the majority of lessons. The National Center for Teacher Residencies (NCTR) advocates for a "clinically oriented teacher preparation" model wherein teams comprised of residents and one or more mentor teachers expose residents to various instructional arrangements, explore data together, and learn to work as part of a professional learning community (National Center for Teacher Residencies, 2016a).

To combine practice with theory, residents simultaneously take graduate coursework to earn a master's degree. There is often a strong emphasis on training residents to serve in culturally diverse school settings (Hammerness & Craig, 2016; Tindle et. al., 2011). Programs typically provide a stipend to residents as they learn to teach in the classroom. These components (i.e., experiential learning, graduate coursework, and a stipend) are the structural backbone of most residency programs.

Residency programs have different perspectives on the role the host teacher plays in professionally developing residents, and they may apply one of several approaches within their program models. Many residency programs use what is called a "mentor" teacher model, wherein host teachers are selected to serve as "mentors" based on their experience, efficacy in the classroom, and willingness to serve as a mentor. Under this model, espoused by NCTR, these carefully selected 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 (National Center for Teacher Residencies, 2015). Mentors are selected by school administrators based on their teaching experience and history of effectiveness. Residency programs typically provide training to host teachers to cultivate them as mentors to residents, along with a stipend in exchange for their efforts. The hypothesis behind the mentor model is that coaching the host teachers and giving them responsibility for the residents will strengthen their investment in and commitment to the resident's training.

Other residency programs use alternative models. For instance, one program consulted as part of this project selects host teachers based only on their willingness to host a resident. In this model, host teachers do not coach or mentor residents. Instead, the residency program provides

necessary the resident with 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 coaching themselves. Another program consulted as part of this project provides residents with classroom experience, but also has them rotate through other positions in the school, including small group tutoring and after school support.

The number of teacher residency programs in the United States has been growing steadily over the last two decades, particularly since Congress created the Teacher Quality Partnership Grants Program to fund and test these programs in new settings (Silva et al., 2014). Given their fairly short history, there is limited research on their impact. Studies mostly focus on the primary or longer-term impact of residency programs on student outcomes. Specifically, researchers have explored whether teachers who graduate from residency programs are more effective teachers by examining performance of students taught by residency graduates. For example, Rosenberg and Miles (2018) assess the return on investment of residency programs by examining impact on student learning, concluding that if a high-need school placed resident graduates from a rigorous teacher residency program in hard-to-staff roles that would otherwise be covered by 1) qualified long-term substitutes or 2) less effective substitutes, students taught by that cohort of resident graduates could gain an average of 3.5 and 4.2 additional months of learning, respectively, in one year. Another study that looks at resident graduates from Boston Teacher Residency uses value-add models and finds that residents become more effective math teachers over time such that by their fourth and fifth years they outperformed teachers who were not resident graduates (Papay et al., 2012). Individual residency programs in Memphis, Denver, San Francisco, and other districts also point to research demonstrating the long-term, positive impact of their programs on teacher

effectiveness (Garrison, 2019; National Center for Teacher Residencies, 2016b). Silva et al. (2014) find that residency program graduates felt more prepared to teach than other new teachers in the same districts, but that retention rates among residency program graduates were comparable to those of other new teachers.

Mechanisms for short-term impact of residency programs

Though research has typically focused on the longer-term benefits of residency programs, hosting a resident could also impact host teacher practice and student learning in the short run. Specifically, hosting another adult in the classroom provides opportunities for team teaching, which in turn may encourage small group instruction and teacher-student relationship building, and may even elevate teaching quality among the host teacher. Team teaching is defined as a classroom teaching model that has "two or more teachers whose primary concern is the sharing of teaching experiences in the classroom, and co-generative dialoguing with each other. They take collective responsibility for maximizing learning to teach or becoming better at teaching while providing enhanced opportunities for their students to learn" (Jang, 2006). Although team-taught classrooms typically have two or more certified teachers who plan for and execute lessons together, other adults in team-taught classrooms are not required to be certified teachers. In this sense, classrooms with a host teacher and resident may be akin to team-taught classrooms, with host teachers and residents co-planning and co-teaching and taking collective responsibility for student learning. Moreover, team teaching allows more experienced teachers to collaborate and work collegially with less experienced teachers and teachers-in-training, providing peer coaching that improves teacher practice and leads to more positive experiences for students (Supovitz, Sirinides & May, 2010).

Team-taught classrooms have been shown to introduce a range of benefits for students. In an exploratory analysis of student survey data, Gladman (2015) finds that students feel team teaching improves their understanding of classroom instruction, increases their willingness to ask questions in class, and enables teachers to take better care of students. In another study comparing outcomes in math among eighth graders in team-taught classrooms versus traditional classrooms, Jang (2006) finds that students in the team-taught group performed significantly better on their final exams. Moreover, students in Jang's study preferred the team teaching model to the traditional model. Johnson et al. (1991) cite benefits of team teaching to students in the form of higher achievement, greater retention, improved interpersonal skills, and an increase in positive interdependence. Likewise, a meta-analysis of quantitative efficacy research by Murawski & Swanson (2001) and a metasynthesis of qualitative research by Scruggs, Mastropieri, & McDuffie (2007) suggest that team teaching has a demonstrated positive effect on student achievement. Administrators, teachers, and students perceived team-teaching to be socially and academically beneficial to all types of students including special education students (Walsh, 2012).

One mechanism by which team teaching may improve student learning is by facilitating small group instruction. 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 enables teachers to more easily facilitate small group learning and help provide individualized, differentiated instruction that is more catered to individual students' needs. Johnson et al.'s (1991) review of over 600 studies comparing the effectiveness of different types of learning experiences finds that cooperative learning environments, which draw on carefully structured small group work to maximize

student learning, produce higher achievement, more positive relationships among students, and healthier psychological adjustment, compared to competitive or individualistic learning. Springer et al.'s (1999) meta-analysis on the effects of small group learning experience on undergraduates in STEM education shows that, relative to traditional classroom settings, small-group learning is effective in promoting academic achievement, more favorable attitudes towards learning, and increased persistence through STEM courses. Hamann et al. (2012) compare students' evaluations of discussions in a political science class that was conducted in three different settings: small discussion groups, whole-class settings, and online discussion group. They find that students identified small discussion groups as promoting the highest satisfaction and critical thinking.

Having an additional adult in the classroom also enables students to forge a relationship with an additional adult, which becomes particularly important if they do not identify closely with the lead teacher but may find that they better connect with the resident. Being able to establish such personal connections can promote ongoing encouragement and support and protect against negative behaviors, further bolstering their attachment to school and quest for learning (O'Connor, Dearing & Collins, 2011; Rudasill et al., 2010).

Finally, to the extent that having an additional adult in the classroom promotes collaboration and professional learning between teachers, it may also lead to more effective teaching practices (Sandholtz, 2000, Smith & Scott, 1990, Gladman, 2015). As teachers co-plan, they may be more likely to locate and share ideas and experiment with new ideas (Robinson & Schaible, 1995). Furthermore, host teachers 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. For instance, a recent Stakeholder Perception Report published by NCTR reports that 98 percent of mentor teachers in their partner programs agreed that participating in a residency program improved their abilities as a teacher leader (National Center for Teacher Residencies, 2018).

To date, little research has examined the short-term impact of hosting a resident on teacher effectiveness or student outcomes in the year the resident is hosted in a classroom. One study by Osborne and Farber (2014) finds that students in classrooms of mentor teachers hosting Tech Teach teacher candidates in Texas perform better on state and district assessment exams across subjects and school levels (i.e., elementary, middle school, high school) compared to students in classrooms of comparable teachers not hosting a teacher candidate. In another study, Papay et al. (2012) apply value-added methods to investigate whether residency programs were able to directly impact student achievement by placing residents in classrooms. They find that Boston Teacher Resident mentors were more effective than non-mentor teachers in raising student achievement in math and English Language Arts (ELA), even after controlling for teacher experience, but there is no evidence that their effectiveness was related to having a resident in the classroom because the mentor's performance in the mentor year or beyond was not statistically different from his or her performance in pre-mentor year.

It remains possible that hosting 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. A study on the effects of having a student teacher in the classroom suggests this can be the case. Using data on

student teaching placements from 14 teacher education programs in Washington State, Goldhaber et al. (2018) find that students perform slightly worse in math and no differently in ELA when there is a student-teacher present in the class, compared to years in which a teacher does not host a student-teacher. However, the researchers find some evidence that more effective host teachers could potentially mitigate any negative impact of hosting a student-teacher on student performance. In this study, the lower math outcome among students in classrooms with student-teachers is driven by host teachers in the lowest quartile of value-added scores, suggesting that more effective host teachers could potentially outweigh the negative impact of hosting a student-teacher on student performance. Furthermore, Ronfeldt et al., (2018) find evidence that effectiveness of the host teacher is positively associated with the future performance of the student teachers when they become teachers.

DATA AND STUDY SAMPLE

To answer the primary research question, we use data from the three residency programs and their respective school partners, described below. To protect the anonymity of the organizations that participated in this study, we use pseudonyms and provide limited information on the background of the organization and their program model. Programs were primarily selected based on their willingness to participate in the study, their scale (programs must have large enough footprints in their partner schools to make a study possible), and their model (programs must pair residents with one host teacher at a time for at least half an academic year but ideally a full year. Two residency programs—Res Ed, and City Teach—use the mentor model espoused by NCTR (i.e., the host teacher serves a mentorship role to the resident). The third program—Teacher Prep—uses the host teacher model (i.e., the host teacher opens their classroom to a resident and the residency program provides mentorship to the resident). 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.

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. We selected partner districts 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 provide background on each of the partner programs and describe the study sample.

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.

Our study draws on data from one of Res Ed's partner CMOs—Charter Plus Academy (Charter Plus). Table 1 summarizes the years for which each partner district provided data. For Res Ed we received data on five cohorts of residents placed with host teachers from 2012–13 to 2016–17. 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 Res Ed host teachers included in the sample.

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

	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

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: 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 schools, 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 have a certain level of effectiveness, as measured by a district-specific teacher effectiveness measure. In addition to the application, City Teach program staff members 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.

City Teach places its residents in both public and charter schools in Regional School District (RSD). Our study includes residents placed with host teachers in both public and charter schools within the RSD district. RSD provided data on five cohorts of residents serving in schools between 2013–14 and 2017–18 (Table 1). 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.[1] 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: 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." The intentional language reflects Teacher Prep's belief that it is difficult to control the pipeline of high-quality host teachers. As such, they seek out teachers who will host a resident for a year, and Teacher Prep takes responsibility for mentoring 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. Teacher Prep provides limited training to host teachers. The training is

[1] 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.).

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, Teacher Prep assigns residents their own coaches, as mentioned above, and offers other opportunities for professional enrichment.

We use 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 (Table 1). After excluding cases that are missing data for baseline or outcome years, the sample includes 101 teachers across the two cohorts (Table 2). 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 [not shown in table]). 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 [not shown in table]). 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 teacher effectiveness scores (TES), defined below under "Dependent variable." We run a multilevel regression model to examine whether the TES values in the outcome year are significantly different for teachers who host residents compared to those who do not, after taking into account TES values in prior years (i.e., TES baseline), years of teaching experience, and teachers' school characteristics. The model includes a binary

variable indicating whether the teacher serves as 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, even if they do not host resident in the subsequent year/s so they are not included in the comparison group.

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 of English Language Learners (ELLs) and percentage of students with Individualized Education Plans (IEPs), 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, $\boldsymbol{\varepsilon}_i$ captures the random error associated with

teacher i and ϵ_j captures the random error associated with school j. We use a mixed effects model to account for multilevel clustering at the teacher level (since there are multiple scores across time for each teacher) and school level (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 this 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 TES values 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 of student achievement 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: entering; emerging; effective; highly effective; or master. Most teachers are rated either effective (28.4 percent) or highly effective (62.2 percent).

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 raises 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). Under this rating scheme, the modal score, with 65.7 percent of teachers sharing this value, is average effectiveness.

Teacher Prep: The TES for Teacher Prep is an "overall performance measure" 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. An explanation of how the points from each goal are then used to determine the overall performance calculations is available upon request. 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. Nearly two-thirds of teachers receive a "meeting high bar" rating.

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 TES values to estimate an average baseline TES for each treatment year included in our analysis. Table 3 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. Therefore, we control for number of years of experience in our model, except in the case of City Teach where data are not available.

Table 3. Academic years included in baseline TES estimates for each cohort, by program.* Source: Program academic datasets.

	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.

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 have IEPs, 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.

Descriptive statistics

Table 4 compares characteristics of host teachers to teachers who did not host residents. For each residency program, we show average characteristics for both groups using the pooled data for all years. 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 of students with IEPs and ELLs.

Table 4. 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. IEP	1.3	1.3	12.0	11.4	7.1	6.6
Pct. ELL	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 characteristic.

FINDINGS

Table 5 summarizes the results of the analysis using the three model specifications, as well as results broken out by year using specification 1. (Appendix A 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 5 present results for each year using specification 1.

Table 5. 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

Res Ed: The results for all years using specification 1 indicate that, controlling for $TES_{baseline}$ and other teacher- and school-level characteristics, Res Ed hosts had an average $TES_{outcome}$ 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 ($p<0.01$). Although both groups' mean TES falls closer to the teacher effectiveness category of "highly effective," being a Res Ed host is associated with an increase in one-fifth of the standard deviation (SD) of $TES_{outcome}$ (the $TES_{outcome}$ has a SD of 0.56 and 0.11/0.56 = 0.20) (see Figure 4a).

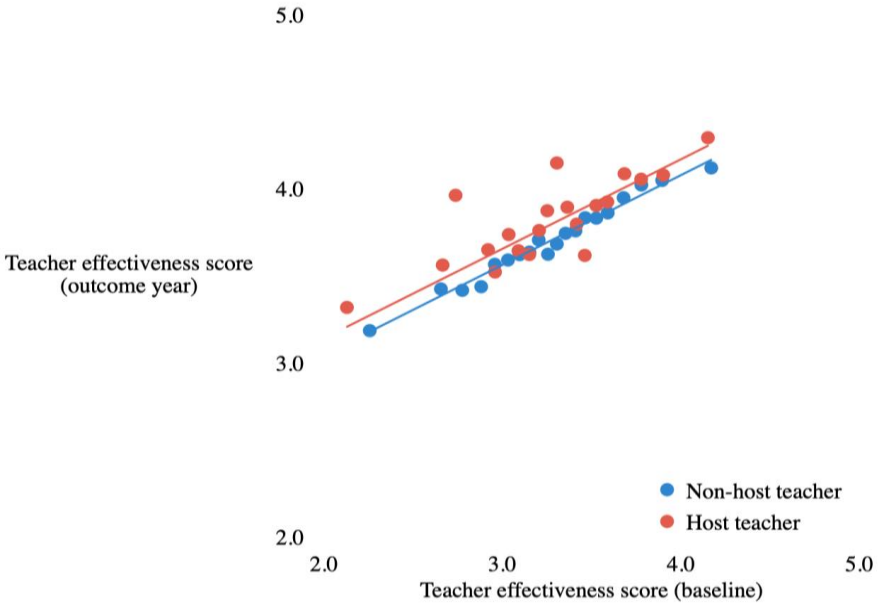
Figure 1 shows the results from specification 1 visually. The graph groups the $TES_{baseline}$ variable into equal-sized bins, and then computes the mean of the $TES_{baseline}$ variable and

$TES_{outcome}$ variables within each bin, to create a scatterplot of these data points (Stepner, 2013). Each dot shows the average $TES_{outcome}$ for a given level of $TES_{baseline}$ for Res Ed hosts (in red) and comparison teachers (in blue). Finally, the plot shows the best linear fit line from a multivariate regression of the $TES_{outcome}$ variable on the $TES_{baseline}$ variable. The graph, therefore, shows the relationship between teachers' $TES_{baseline}$ and their $TES_{outcome}$, so that we have a sense of host and comparison teachers' average TES in the outcome year, given their baseline TES. 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 indicate that host teachers' $TES_{outcome}$ 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. 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 analyses separately by year, there is no significant difference between the outcome $TES_{outcome}$ for Res Ed hosts versus other teachers in any year except for academic year 2012–13: in this year, being a Res Ed host is associated with an increase in $TES_{outcome}$ of 0.21, and the difference is statistically significant ($p<0.01$).

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 $TES_{baseline}$ and other teacher- and school-level characteristics, City Teach hosts had an average $TES_{outcome}$ 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 ($p<0.01$).[2] 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 SD of $TES_{outcome}$ (the outcome TES has a SD of 1.9 and 0.6/1.9 = 0.32) (see Figure 4b).

Figure 2a shows the City Teach results from specification 1 visually. In the graph, we see more red dots above blue dots than below; although it appears that there are some outliers at the

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

lowest and highest ends of the distribution that could be affecting the mean estimates. To confirm that they are not biasing the results, we remove all outliers with $TES_{baseline}$ 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, the red host curve is more consistently above the blue curve.

In specification 2, City Teach host teachers outperformed the non-hosts by 0.78 standard error units, an increase in 0.41 of the SD of $TES_{outcome}$ (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 $TES_{outcome}$ 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 $TES_{outcome}$ 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 $TES_{outcome}$ of 0.54 standard error units, but the difference is only marginally statistically significant ($p<.10$).

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.

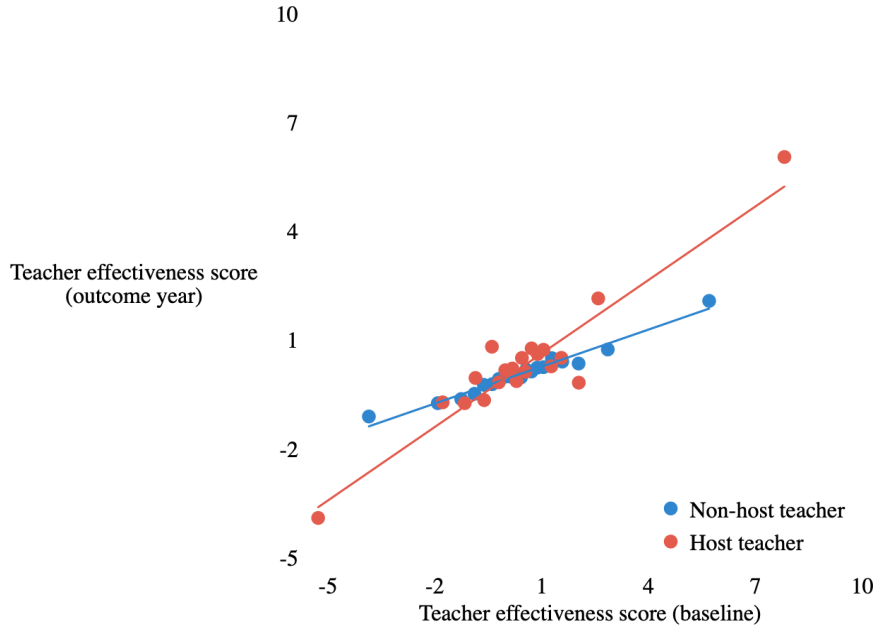
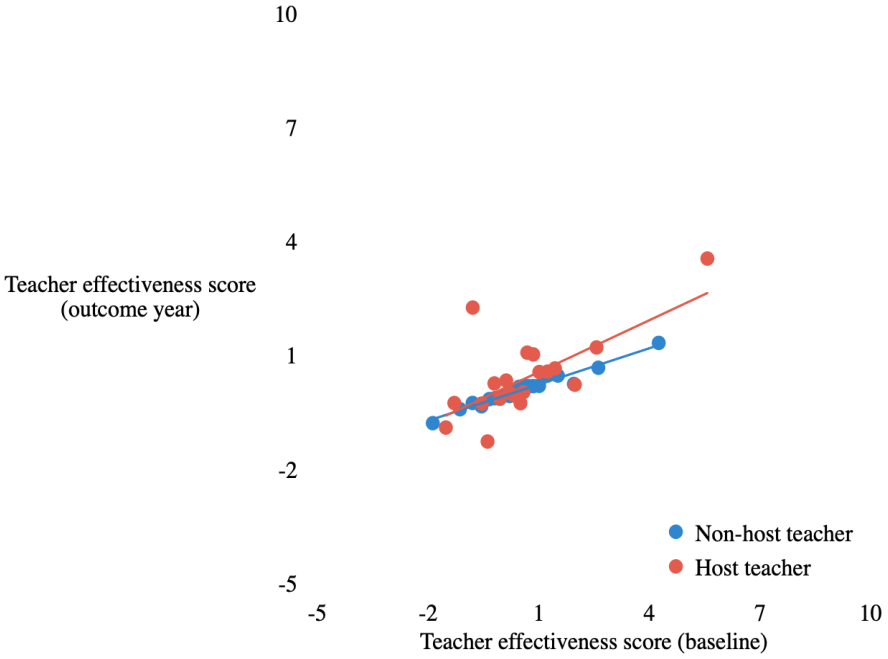


Figure 2b. Relationship between teacher effectiveness score (baseline) and teacher effectiveness score (outcome year) for City Teach, based on model specification 1, without outliers. Source: Residency and partner district administrative records.

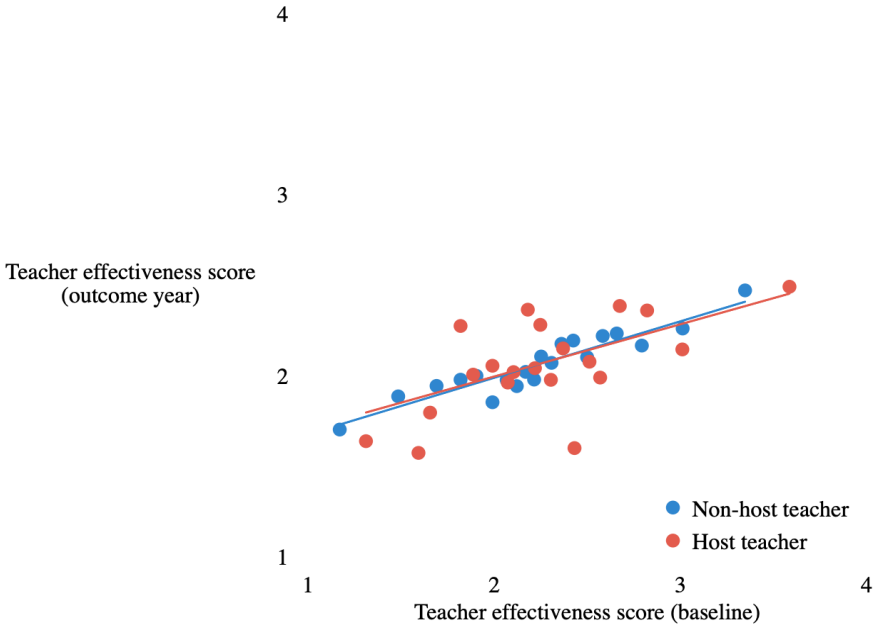


Teacher Prep: The results for all years using specification 1 indicate that the difference between host and comparison teachers for all years was small, negative (-.02), and statistically insignificant. On average, controlling for $TES_{baseline}$ and school-level characteristics, Teacher Prep hosts had an average $TES_{outcome}$ of 2.02, compared to 2.04 among teachers who did not host a resident. Although both groups' mean TES values 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 SD of $TES_{outcome}$ (the $TES_{outcome}$ has a SD of 0.59 so 0.02/0.59=0.03) (see Figure 4c). Figure 3 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.[3]

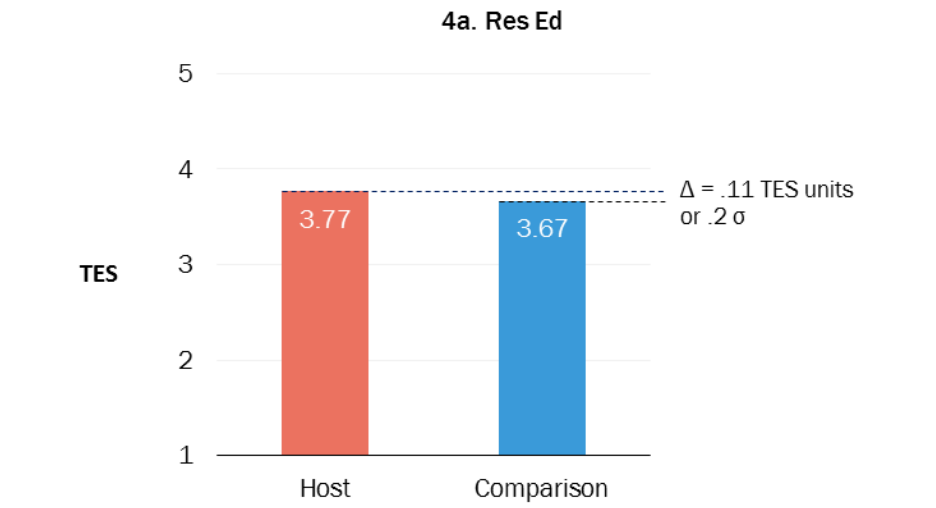
The specification 2 and 3 results do not change much from specification 1. When we conduct the analyses separately by year, there is no significant difference between the $TES_{outcome}$ for Teacher Prep hosts versus other teachers in either year.

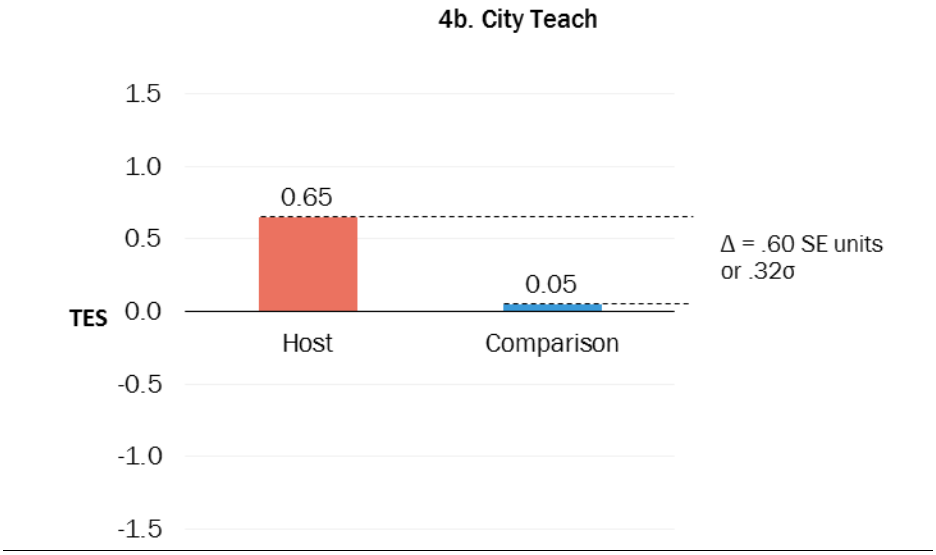
[3] Because the Teacher Prep TES only includes three categories, as a robustness check, we also ran an ordered logistic regression to examine whether being a Teacher Prep host teacher changes the likelihood of receiving a higher rating. The results indicate there is no effect. We similarly repeated the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

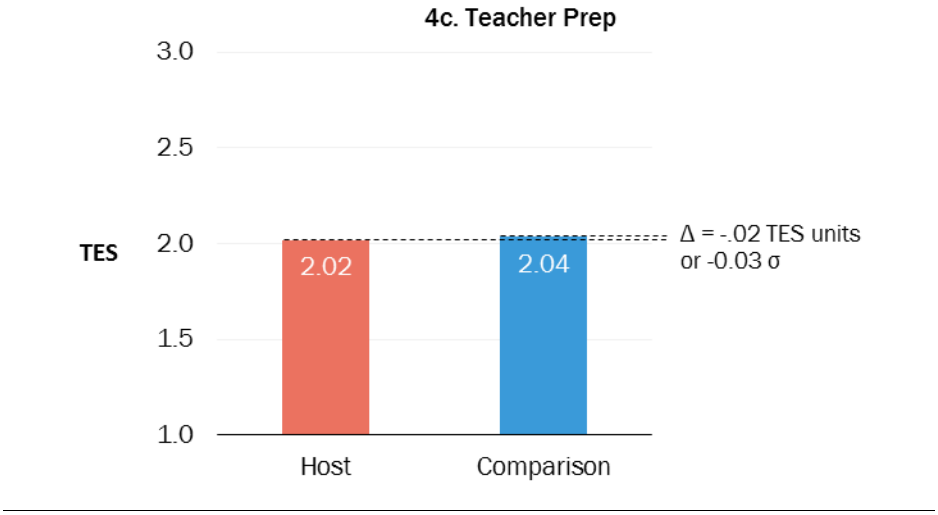
Figure 3. Relationship between teacher effectiveness score (baseline) and teacher effectiveness score (outcome year) for Teacher Prep, based on model specification 1. Source: Residency and partner district administrative records.



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







DISCUSSION

Based on the results for Res Ed and City Teach, we find evidence that hosting a resident in the classroom may be associated with higher TES values 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 a major factor (and the only factor in the case of City Teach) contributing to the calculation of a teacher's TES, higher TES values 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 TES values 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. Teacher Prep does not emphasize teacher experience or quality when selecting host teachers. On the other hand, Res Ed and City teach follow the mentor model and, therefore, have stricter selection criteria for host teachers. Our descriptive statistics confirm that, consistent with the program model, Res Ed and City Teach are recruiting better qualified and more experienced teachers based on average TES values and years of teaching experience. 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 TES values 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 an interaction effect in the case of Res Ed. However, we did find a significant interaction effect between being a City Teach host and the $TES_{baseline}$ (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 $TES_{baseline}$. Hosts with higher $TES_{baseline}$ had significantly higher $TES_{outcome}$, where each unit increase in $TES_{baseline}$ is associated with a 0.36 standard error unit increase in their $TES_{outcome}$. Thus, it is plausible that stronger host teachers in this program are better able to translate having a resident in the classroom to improved outcomes.

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. 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 A3 of Appendix A).

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 study. 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. 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 available upon request.) Although the sample sizes are small, these results are a springboard for further study on the underlying mechanisms of the effect.

SUMMARY

The paper provides evidence that, in some settings, there are ancillary benefits of residency programs. We explore whether hosting a resident in the classroom improves teacher effectiveness, as measured by TES values. The results for Res Ed and City Teach are consistent in that we find a significant, positive effect for both programs that share a similar model, where hosting a resident in the classroom is associated with improvements in TES values of teachers and hence, improved student learning. We did not find a significant effect of hosting a resident in the classroom for Teacher Prep. This finding is interesting, and in many ways logical, because Teacher Prep uses a different program model than Res Ed and City Teach. As described earlier, Res Ed and City Teach use the "mentor teacher" 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. Teacher Prep, on the other hand, uses the non-mentor 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 4. Teacher Prep teachers hosting residents have roughly the same experience and prior effectiveness scores as Teacher Prep teachers who do not host residents, whereas host teachers at Res Ed and City Teach are generally more experienced and have higher prior effectiveness scores than the comparison group. 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 interaction terms between being a host and

baseline TES, as well as being a host and teaching experience for each program, and found a significant interaction effect only in the case of City Teach, suggesting that City Teach hosts with higher baseline TES had significantly higher outcome TES.

It is worth noting that we found no evidence that hosting a resident in a classroom was associated with lower teacher effectiveness score 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 Teacher Prep) 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 only find evidence that Teacher Prep 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.

Moving forward, researchers can build on this work in two primary ways. First, this research was based on three residency programs working in three school districts. There are clear opportunities for replication, with the goal of including data from not only from additional residency programs but also residency programs that may use different models to select host teachers and train and support both residents and host teachers. Additionally, we cannot yet

answer the question of why hosting a resident in the classroom may have benefits for teacher effectiveness and student learning. This paper describes a number of hypotheses, and additional research should focus on underlying mechanisms. Teacher residency programs will benefit from a better understanding of the types of teacher characteristics that may mediate the effectiveness of the teacher-resident partnership; the types of host teacher training and support that yield effective partnerships; the types of training that maximizes residents' short-term impact in the classroom; and the ways in which residency programs can nurture strong relationships between residents and host teachers.

References

Arhar, J. M., & Walker, D. (2002). A collaborative and developmental approach to student teaching in the middle level. *Middle School Journal*, *33*(5), 24-32.

Backes, Ben, and Michael Hansen. (2018). Reaching Further and Learning More? Evaluating Public Impact's Opportunity Culture Initiative. National Center for Analysis of Longitudinal Data in Education Research, Working Paper 181.

Cook, L., and M. Friend. (1996). Coteaching: guidelines for creating effective practices, in: E. L. Meyen, G. A. Vergason & R. J. Whelan (Eds) Strategies for teaching exceptional children in inclusive settings (Denver, OH, Love). 155 – 182.

Gladman, A. (2015). Team teaching is not just for teachers! Student perspectives on the collaborative classroom. TESOL Journal, 6(1), 130-148.

Garrison, A.W. (2019). Memphis Teacher Residency: Teacher Effectiveness in 2018–19. Department of Research and Performance Management, Shelby County Schools.

Goldhaber, D., Krieg, J. M., & Theobald, R. (2017). Does the match matter? Exploring whether student teaching experiences affect teacher effectiveness. *American Educational Research Journal*, *54*(2), 325-359.

Goldhaber, Dan, John Krieg, Roddy Theobald. (2018). The Costs of Mentorship? Exploring Student Teaching Placements and Their Impact on Student Achievement. CALDER Working Paper No. 187, March 2018

Guha, R., Hyler, M. E., & Darling-Hammond, L. (2017). The teacher residency: A practical path to recruitment and retention. *American Educator*, *41*(1), 31.

Hammerness, K., & Craig, E. (2016). "Context-Specific" Teacher Preparation for New York City: An Exploration of the Content of Context in Bard College's Urban Teacher Residency Program. *Urban Education*, *51*(10), 1226-1258.

Jang, S. J. (2006). Research on the effects of team teaching upon two secondary school teachers. *Educational research*, *48*(2), 177-194.

Johnson, D.W., R.T. Johnson, and K.A. Smith. (1991). Cooperative Learning Increasing College Faculty Instructional Productivity. *ASHE-ENC Higher Education Report No. 4*. Washington, D.C.: The George Washington University, School of Education and Human Development. Hamann, Kerstin, Philip H. Pollock, and Bruce M. Wilson. (2012). Assessing Student Perceptions of the Benefits of Discussions in Small-Group, Large-Class, and Online Learning Contexts, College Teaching, 60:2, 65-75, DOI: 10.1080/87567555.2011.633407

Mueller, S. (2013). Teacher experience and the class size effect—Experimental evidence. *Journal of Public Economics*, *98*, 44-52.

Murawski, W. W., and H. Lee Swanson. (2001). A meta-analysis of co-teaching research: Where are the data?. *Remedial and special education*, *22*(5), 258-267.

National Center for Teacher Residencies. (2015). *Clinically oriented teacher preparation.* National Center for Teacher Residencies.

National Center for Teacher Residencies. (2016a). *The Case study project: Clinically oriented teacher preparation.* National Center for Teacher Residencies.

National Center for Teacher Residencies. (2016b). 2015 Network Impact Overview; Research Brief. National Center for Teacher Residencies.

National Center for Teacher Residencies. (2018). *2017 Stakeholder Perception Report*. National Center for Teacher Residencies.

O'Connor, E. E., Dearing, E., & Collins, B. A. (2011). Teacher-child relationship and behavior problem trajectories in elementary school. *American Educational Research Journal*, *48*(1), 120-162.

Osborne, C., and M. Farber. (2014). Were the promises of TechTeach realized? Evaluation report for Part A (Teacher Preparation) of the i3 grant at Texas Tech University. Center for Health and Social Policy, The LBJ School of Public Affairs, The University of Texas at Austin.

Papay, J. P., West, M. R., Fullerton, J. B., & Kane, T. J. (2012). Does an urban teacher residency increase student achievement? Early evidence from Boston. *Educational Evaluation and Policy Analysis*, *34*(4), 413-434.

Robinson, B., and R. M. Schaible. (1995). Collaborative teaching: Reaping the benefits. *College Teaching*, *43*(2), 57-59.

Rosenberg, D., & Miles, K. H. (2018). Growing Great Teachers: How School System Leaders Can Use Existing Resources to Better Develop, Support, and Retain New Teachers--and Improve Student Outcomes. *Education Resource Strategies*.

Rudasill, K. M., Reio, Jr., T. G., Stipanovic, N., & Taylor, J. E. (2010). A longitudinal study of student–teacher relationship quality, difficult temper-ament, and risky behavior from childhood to early adolescence. Journal of School Psychology, 48, 389–412.

Sandholtz, J. H. (2000). Interdisciplinary Team Teaching as a Form of Professional Development. *Teacher Education Quarterly,* Vol. 27, No. 3, Critical Analysis and Reflective Practice, pp. 39-54.

Scruggs, T. E., M. A. Mastropieri, and K. A. McDuffie. (2007). Co-teaching in inclusive classrooms: A metasynthesis of qualitative research. *Exceptional Children, 73*(4), 392-416.

Silva, T., McKie, A., Knechtel, V., Gleason, P., & Makowsky, L. (2014). *Teaching residency programs: A multisite look at a new model to prepare teachers for high-need schools* (No. 58dae816db6b4d008b73ac09df6a92cb). Mathematica Policy Research.

Smith, S., and J. Scott. (1990). The collaborative school: A work environment for effective instruction. Eugene, OR: ERIC Clearinghouse on Educational Management.

Solomon, J. (2009). The Boston teacher residency: District-based teacher education. *Journal of Teacher Education, 60*(5), 478-488.

Springer, L., M. E. Stanne, and S. S. Donovan. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of educational research, 69*(1), 21-51.

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

Supovitz, J., Sirinides, P., & May, H. (2010). How principals and peers influence teaching and learning. *Educational Administration Quarterly*, *46*(1), 31-56.

Tindle, K., Freund, M., Belknap, B., Green, C., & Shotel, J. (2011). The urban teacher residency program: A recursive process to develop professional dispositions, knowledge, and skills of candidates to teach diverse students. *Educational Considerations, 38*(2), 28-35.

Walsh, J. M. (2012). Co-teaching as a school system strategy for continuous improvement. *Preventing School Failure: Alternative Education for Children and Youth*, *56*(1), 29-36.

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

Table A1. 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)
Academic year (ref: 2013 – 14)						
2014 – 15	0.32***	--	--	--	--	--
	(0.053)					
2015 – 16	-0.05	--	--	--	--	--
	(0.076)					
2016 – 17	-0.38***	--	--	--	--	--
	(0.078)					
2017 – 18	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 A2. 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)
Academic year (ref: 2012 – 13)						
2013 – 14	-0.09**	--	--	--	--	--
	(0.036)					
2014 – 15	0.07*	--	--	--	--	--
	(0.037)					
2015 – 16	0.03	--	--	--	--	--
	(0.037)					
2016 – 17	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 A3. Teacher Prep

VARIABLES	All years	2016–17 Only	2017–18 Only	All years
Teacher Prep host	-0.02	-0.04	-0.01	--
	(0.055)	(0.073)	(0.078)	
No. months worked as Teacher Prep 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)
Academic year (ref: 2015 – 16)				
2016 – 17	-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