NCEE 2012-4051

U.S. DEPARTMENT OF EDUCATION

Moving Teachers: Implementation of Transfer Incentives in Seven Districts

NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences

Moving Teachers: Implementation of Transfer Incentives in Seven Districts

April 2012

Steven Glazerman
Ali Protik
Bing-ru Teh
Julie Bruch
Neil Seftor
Mathematica Policy Research

Elizabeth Warner
Project Officer
Institute of Education Sciences

NCEE 2012-4051
U.S. DEPARTMENT OF EDUCATION

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences

U.S. Department of Education
Arne Duncan
Secretary

Institute of Education Sciences
John Q. Easton
Director

National Center for Education Evaluation and Regional Assistance
Rebecca A. Maynard
Commissioner

April 2012

The report was prepared for the Institute of Education Sciences under Contract No. ED-04-CO-0112/007. The project officer is Elizabeth Warner in the National Center for Education Evaluation and Regional Assistance.

Glazerman, S., A. Protik, B. Teh, J. Bruch, N. Seftor. (2012). *Moving Teachers: Implementation of Transfer Incentives in Seven Districts* (NCEE 2012-4051). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

To order copies of this report:

- Write to ED Pubs, Education Publications Center, U.S. Department of Education, P.O. Box 1398, Jessup, MD 20794-1398.
- Call in your request toll free to 1-877-4ED-Pubs. If 877 service is not yet available in your area, call 800-872-5327 (800-USA-LEARN). Those who use a telecommunications device for the deaf (TDD) or a teletypewriter (TTY) should call 800-437-0833.
- Fax your request to 301-470-1244.
- Order online at www.edpubs.org.

This report also is available on the IES website at http://ies.ed.gov/ncee.

Upon request, this report is available in alternate formats such as Braille, large print, audiotape, or computer diskette. For more information, please contact the Department's Alternate Format Center at 202-260-9895 or 202-205-8113.

ACKNOWLEDGMENTS

This study is the product of many people's efforts. Unfortunately we cannot acknowledge them by name, but we are deeply grateful to the many teachers, principals, district leaders, and central office staff whose hard work and patience made both the intervention and the research possible. Also spearheading the implementation of the intervention were staff from The New Teacher Project (TNTP), including Coral Jenrette, Mónica Vásquez, Emma Cartwright, Latricia Barksdale, and Kristen Rasmussen. At Mathematica, Tim Silva and Jeffrey Max played important roles in overseeing implementation and working closely with TNTP and the districts. Monica Leal Priddy at Optimal Solutions Group led a team, including Kimberly Hahn, Carolina Herrera, Grace Hong, Theodore Hickey, and Sarah Newman, that also played an important role in gathering data needed for program implementation.

A large data collection effort was necessary to make the study successful. This report relies most heavily on teacher and principal surveys. At Mathematica, Nancy Carey and Kristina Rall led the survey effort with invaluable assistance from Theresa Boujada and her team at Mathematica's Survey Operation Center.

The evaluation team at Mathematica benefited from expert programming and research assistance from Norma Altshuler, Alena Davidoff-Gore, Maureen Higgins, Christopher Jones, Jessica Peterson, and Claire Smither Wulsin. John Deke and Philip Gleason read and provided helpful comments on earlier versions of the report. Jeffrey Max played a major role in random assignment. A Technical Work Group (TWG) provided useful input on program design and the research. TWG members included Dale Ballou, Jason Kamras, Robert Meyer, Anthony Milanowski, Jeffrey Smith, and Jacob Vigdor. The report was edited by Betty Teller and prepared for publication by Donna Dorsey.

This page has been left blank for double-sided copying.

DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST[1]

The research team for this evaluation consists of a prime contractor, Mathematica Policy Research of Princeton, New Jersey, and a subcontractor, Optimal Solutions Group of College Park, Maryland. Neither of these organizations or their key staff members have financial interests that could be affected by findings from the evaluation. No one on the Technical Working Group, convened by the research team to provide advice and guidance, has financial interests that could be affected by findings from the evaluation.

[1] Contractors carrying out research and evaluation projects for IES frequently need to obtain expert advice and technical assistance from individuals and entities whose other professional work may not be entirely independent of or separable from the tasks they are carrying out for the IES contractor. Contractors endeavor not to put such individuals or entities in positions in which they could bias the analysis and reporting of results, and their potential conflicts of interest are disclosed.

This page has been left blank for double-sided copying.

CONTENTS

III (*continued*)

TABLES

FIGURES

I. INTRODUCTION

This report describes the implementation and intermediate impacts of an intervention designed to provide incentives to induce a school district's highest-performing teachers to work in its lowest-achieving schools to improve student achievement. The report is part of a larger study that used random assignment to form equivalent groups of classrooms ("teacher teams") that either had the chance to participate in the intervention or did not. A future report will focus on the impacts of the intervention on student achievement and other outcomes.

A. Policy Problem: Unequal Access to Top Teachers

There is growing concern that the nation's most effective teachers are not working in the schools with the most disadvantaged students (Goldhaber 2008; Peske and Haycock 2006). Much of this concern is motivated by research examining the disparity in teacher characteristics such as experience or certification, which are viewed as proxies for teacher effectiveness (Presley et al. 2005; Lankford et al. 2002; Education Trust 2008; Clotfelter et al. 2006; Carroll et al. 2000). These studies show that schools serving a high proportion of low-income or minority students are more likely to employ novice teachers and teachers who lack certification in their subject area than schools serving fewer disadvantaged students.

Teacher characteristics and teacher effectiveness, however, are not equivalent. Because the link between these teacher characteristics and student achievement has not been well established (Rivkin et al. 2005; Gordon et al. 2006; Rockoff et al. 2008; Buddin and Zamarro 2008), more recent analysis has focused on teacher effectiveness in the classroom as measured by student achievement growth, and how teachers identified as effective along this dimension are distributed across schools with higher and lower proportions of disadvantaged students. These measures are referred to as value-added estimates, because they seek to describe the contribution that teachers make (value that they add) to student achievement growth, holding constant factors outside the teacher's control, such as student background and prior learning (McCaffrey et al. 2004; Lipscomb et al. 2010). The question then becomes whether high value-added teachers are underrepresented in schools serving disadvantaged students, as measured in terms of poverty, low achievement, or other factors.

Recent research using teacher value-added measures also raises concerns about access to the highest-performing teachers for disadvantaged students. Three recent studies estimated teacher effectiveness in terms of student achievement growth and examined its distribution in low- and high-poverty schools. Each of the studies used value-added measures of teacher performance, which assume that any differences remaining in average achievement after accounting for prior student achievement and student demographic characteristics are due to the teacher. One study used data from Tennessee and showed that schools with higher percentages of low-income and minority students had fewer of the most effective teachers and more of the least effective teachers on the state's value-added assessment (Tennessee Department of Education 2007). Another study using data from North Carolina and Florida found that the average teacher effectiveness was lower in high-poverty schools, but the prevalence of the most effective teachers was similar across poverty levels (Sass et al. 2010). The third study was based on data that largely overlap with the sample used in the current study. Glazerman and Max (2011) examined the prevalence of districts' highest-performing teachers (in terms of value added) in

elementary and middle schools in 10 large and diverse school districts.[2] The paper demonstrates that, on average, schools with the most disadvantaged students had a significantly lower percentage of teachers who were in the top 20 percent of the performance distribution as measured by value added. This was true whether student disadvantage was measured by income or prior achievement at the middle school level; it was significant at the elementary school level only when prior achievement was used as the measure of student disadvantage.

Policymakers at the federal, state, and local levels have considered a range of options for helping struggling schools attract and retain effective teachers. One goal of such policies is to improve the access that disadvantaged students have to top teachers. The strategies include alternative teacher preparation and certification programs, recruitment bonuses for serving in hard-to-staff schools or subjects, intensive mentoring and professional development, and performance-based pay. The strategies have been implemented with federal funds and with funding from state, local, and nongovernment sources. Some of these policies have been implemented in the context of research studies to gauge their effectiveness, but to date there is little rigorous evidence of any one of these strategies demonstrating clear success in raising student achievement in the U.S.

B. One Policy Response: Transfer Incentives for Highest-Performing Teachers

One strategy that has not been studied in sufficient detail is the use of monetary recruitment incentives that are targeted specifically to teachers with high value-added performance. The U.S. Department of Education's Institute of Education Sciences (IES) has sponsored a research study, summarized in this paper and in a future report, which tests the effectiveness of an intervention that adopts this strategy. The intervention, known to participating school districts as the Talent Transfer Initiative (TTI), identifies a district's highest-performing teachers using value-added analysis and offers them a monetary incentive to transfer to any one of the district's low-achieving schools targeted for the intervention.

IES contracted with Mathematica Policy Research to design an intervention that uses this strategy, oversee its implementation, and conduct a rigorous evaluation of its impact. The implementation itself was carried out in collaboration with the participating school districts by Mathematica's subcontractor, The New Teacher Project (TNTP). The program design and study design were reviewed by a technical working group of experts in the fields of teacher compensation, value-added analysis, and program evaluation. Once the broad parameters of the intervention were defined, such as the method of identifying high-performing teachers and the size and form of the incentives, TNTP developed most of the operational details including the timeline, school and teacher recruitment strategies, and communication plan in consultation with participating school districts. The resulting intervention design is described next.

[2] A "large and diverse" district was defined as one having more than 40 elementary schools, of which at least 10 had a "low" percentage, defined as 40 percent or fewer, of students eligible for free or reduced-price lunch (FRL) and at least 15 had a "high" percentage of FRL-eligible students, defined as 70 percent or more.

1. Overview of the Talent Transfer Incentive

The intervention was designed to proceed as follows. The first step is to conduct value-added analysis of student test scores to identify the highest-performing teachers, defined as the top 20 percent, in each district in the tested grades and subjects based on a value-added measure.[3] The TTI relies on at least two years (and typically three years, depending on district data availability) of student achievement growth data for each teacher, where the data for each year consist of a post-test from the end of the current school year and a pre-test from the end of the previous school year. This analysis is typically completed between January and March of the calendar year in which implementation begins. This corresponds to the school year just before transfer teachers actually would start in their new schools. The amount of time required for the value-added analysis can vary greatly depending on the availability and quality of data on student test scores, demographics, enrollment, and student-teacher links.

Once the highest-performing teachers are identified, the program's second step is to offer them a series of payments, totaling $20,000 over two years, for transferring into and remaining in one of the targeted low-achieving schools within their district. "Low-achieving" schools (those with low average test scores) are targeted because it is a measure of disadvantage that high-performing teachers are in the best position to address.[4] Once the district approves the list, the offer is made in the spring by inviting the highest-performing teachers (referred to as transfer candidates) to an information session where they are recognized for their past accomplishments and asked to consider applying to transfer as a way to help disadvantaged students, who may benefit the most from their talent. Within each district, a site manager follows up with transfer candidates individually to encourage them to apply for a transfer position as part of TTI. The site managers in the TTI study were employees of TNTP who coordinated their outreach activities with each district's human resources department. At the same time that the site manager is recruiting transfer candidates, he or she also works with potential receiving-school principals, beginning with hosting an information session and tracking teaching positions that are expected to be vacant in those schools for the coming fall. Once eligible vacancies are identified, confirmed, and assigned to the program, the site manager performs a matchmaking function, assisting both receiving-school principals and transfer candidates by setting up interviews.

Applicants must interview with and be accepted by the principal at the receiving school and then voluntarily transfer in order to qualify for the additional compensation. Ideally, these offers and acceptances are finalized by early summer.

Finally, the teachers who transfer are provided with an orientation by TNTP in each district and the first installment of their bonus ($2,500) just before the start of the school year. Because

[3] As mentioned above, value-added analysis is the statistical approach that tries to determine the unique contribution each teacher makes to student achievement, holding constant factors that are outside the teacher's control. The specific value-added model used in this study is discussed further in Appendix A.

[4] To define a low-achieving school, all schools were ranked by their average test scores and the TTI team worked with the district administrators to identify schools among the lowest scoring that were not already participating in a program that was similar to TTI. Targeted grades and subjects, typically multiple subject classrooms in grades three through five and math and English language arts classrooms in grades six through eight, were those in which standardized tests were administered in the current and prior grade.

the transfer teachers have proven themselves in the classroom, they typically do not require additional formal support, but a staff person is assigned to provide informal support and answer questions as needed during the first two school years. Over that time, the teachers who remain in their originally assigned positions receive payments at the end of each semester (December and June).

Not all of the districts' highest-performing teachers will have been teaching in the potential sending schools. The transfer incentive is predicated on the idea that low-achieving schools need more high-performing teachers, but some of the most effective teachers may in fact already be working in the potential receiving schools. A critical component of the transfer incentive program is a provision whereby teachers in the highest-performing group who are already teaching in low-achieving schools automatically qualify for a retention bonus of $10,000. This bonus is also paid in installments over two years, as long as the qualifying teachers remain in their schools.

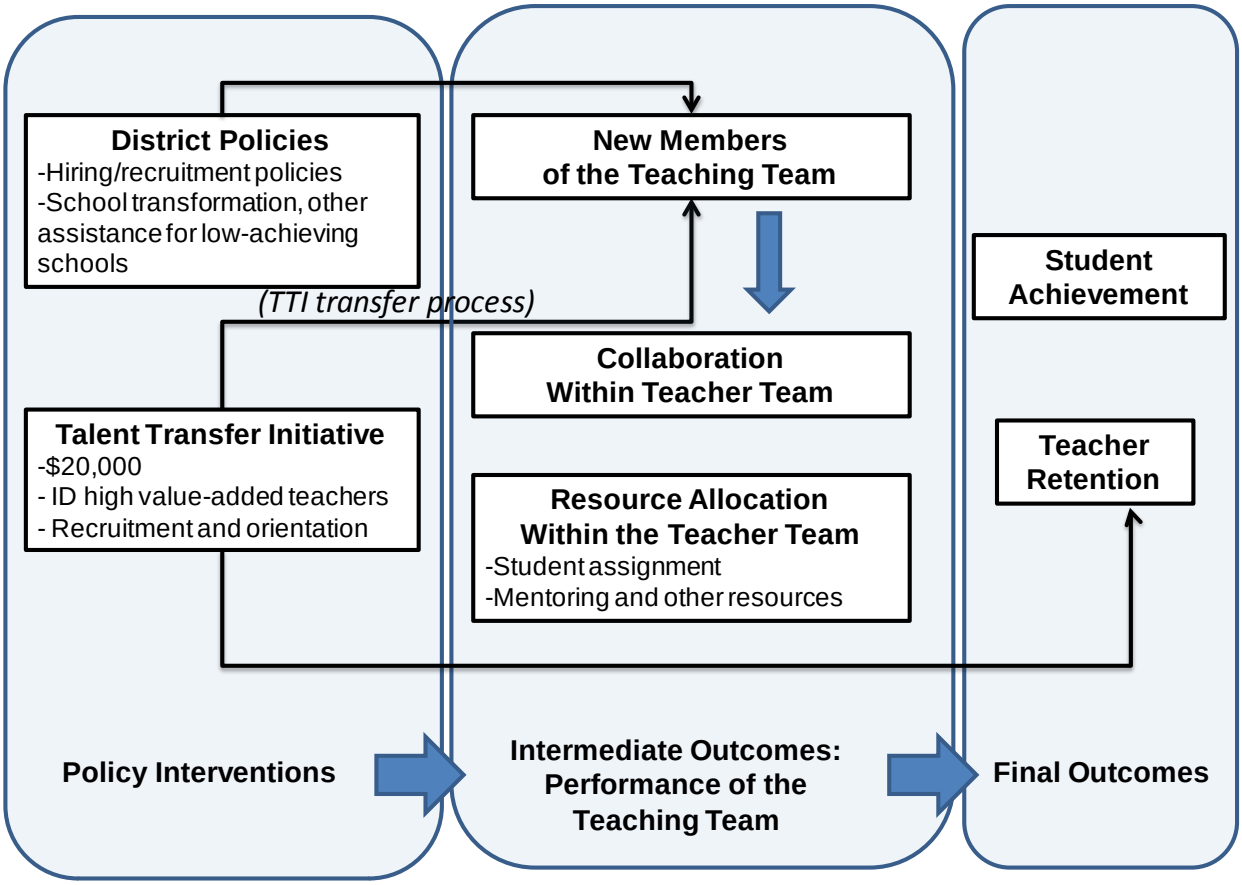
2. Logic Model: How Can Teacher Transfer Incentives Raise Student Achievement?

We hypothesize that a teacher transfer intervention like TTI would improve teacher retention and student achievement outcomes through a series of pathways depicted in a logic model in Figure I.1.

First, TTI would operate in the context of a set of existing district policies aimed at helping low-achieving schools with teaching vacancies to raise teacher performance. The existing policies can include transfer rules, transfer incentives, signing bonuses, or other recruitment and hiring strategies. They can include a range of more general school improvement strategies, such as school turnaround, class size reduction, curricular changes, or changes to the working environment, such as increasing teacher induction and mentoring or hiring a new principal.

Into this context, TTI represents a new intervention that uses the unique tools of identifying the highest-performing teachers using a value-added model and offering substantial monetary incentives ($20,000) to encourage the identified teachers to transfer. We hypothesize that several components of the transfer incentive intervention will influence the probability of a high-performing teacher transferring. These include the criteria used to identify transfer candidates (the value-added model and the cutoff), the size of the incentive, and how the incentive is offered (for example, with a concerted recruiting effort that appeals to candidates' sense of duty). These program factors will combine with other factors such as the attributes that make sending schools more or less desirable as places to remain, attributes that make the receiving schools more or less desirable to transfer in, the match between the teacher and the principals in the sending and receiving schools, as well as the relative desirability of commuting to the sending versus receiving school. We hypothesize that if a high-performing teacher does not fill an identified vacant position, then that vacancy may be filled by a teacher who is new to the school, the district, or the profession, or a teacher who moved from within the school. Another possibility is that the vacancy is lost because student enrollment declines or the teacher who had planned to leave to create the vacancy changes his or her plans.

Figure I.1. Logic Model: How Transfer Incentives Affect Teachers and Students



It is important to note that TTI is voluntary, not mandatory. That means that we do not seek to estimate the impact of a particular teacher forced into a particular school setting. Instead, we wish to estimate the impact of *offering* a school the opportunity to hire from a pool of candidates identified in a particular way (using value-added analysis). This effort amounts to estimating the impact of whichever eligible teachers, if any, happened to apply, successfully transfer into, and remain in targeted teaching positions. A careful recruitment strategy with match-making assistance provided to principals and an orientation for new transfers are meant to facilitate more successful transfers.

The principle way that TTI would have an impact on final outcomes such as student achievement is through intermediate outcomes, specifically, by improving the performance of the teaching team that is targeted by the intervention (one that begins with a teaching vacancy to be filled). The transfer incentive most directly affects the quality of the person filling the vacancy on such a team. We refer to that teacher's impact on his or her students as the direct impact of the intervention. However, we consider all teachers in the team, because whoever fills the vacancy can have indirect effects on students and teachers in the same school and grade. (At the middle school level, we think of a teaching team as anyone who teaches a particular grade and subject at that school, for example, those who teach grade 7 language arts at a given school.

One type of indirect effect operates through collaboration within the teacher team. An experienced, higher-performing transfer teacher might, for instance, help a junior colleague improve lesson planning. Research on student achievement gains in North Carolina suggests that such teacher peer effects can be substantial and lasting (Jackson and Bruegmann 2009).

Another type of indirect effect would operate through the way that resources are allocated within the teaching team. These resources can include mentoring and coaching, for example. If TTI were to result in an experienced, accomplished teacher filling a vacancy in a hard-to-staff school, then a literacy coach in the school might have more time to spend with the existing teachers in the team than he or she would have had if the new member had been new to the profession.

Another important indirect effect within the school is that principals could assign students to teachers differently if they have a different type of teacher filling the vacancy. We do not normally think of students as "resources", but this is another way in which workload, represented by students with greater academic or behavioral challenges for the teacher, might be allocated differently in response to a transfer incentive.

We will attempt to measure the characteristics of new members, the nature of collaboration, and the extent of resource re-allocation within the teacher team, but our final goal is to estimate the net effect of these direct, indirect, and resource allocation effects, which we refer to as the "total effect" on the teacher team. This is captured by the impact on student achievement as measured by standardized test scores.

The other final outcome to measure is teacher retention. A transfer incentive initiative like TTI would have its most direct impact on retention through the phase-in of stipend payments. That is, the $20,000 incentive is paid out in installments over a two-year period, encouraging teachers to stay in order to collect payments. The other way that a transfer incentive will affect teacher retention is by improving satisfaction with and attachment to the principal and school for all teachers in the team. Therefore it is important to measure retention during the two-year commitment period over which the payments are made as well as after that period.

It is worthwhile considering weak links in the causal chain that might hinder success of an intervention like TTI. The main consideration is the quality of the teachers identified as transfer candidates. The two assumptions underlying the transfer incentive intervention are (1) that teachers who have been identified as highest performing will continue in future years to generate large learning gains and (2) that they can continue to be effective in their new settings after they transfer, particularly in contrast with teachers who would ordinarily be hired by low-achieving schools. However, any measure of performance would be an estimate based on current or recent teaching wherever that teacher had been assigned. This makes it especially challenging to know whether a strong estimated performance in the years leading up to the determination of a teacher's status as a highest-performing teacher will predict performance in a following year, in another setting, or with new students. The mechanism by which a transfer strategy like this might succeed or fail at raising student achievement becomes apparent if we express a teacher's measured performance in a given time period as the sum of several parts (see Figure I.2).

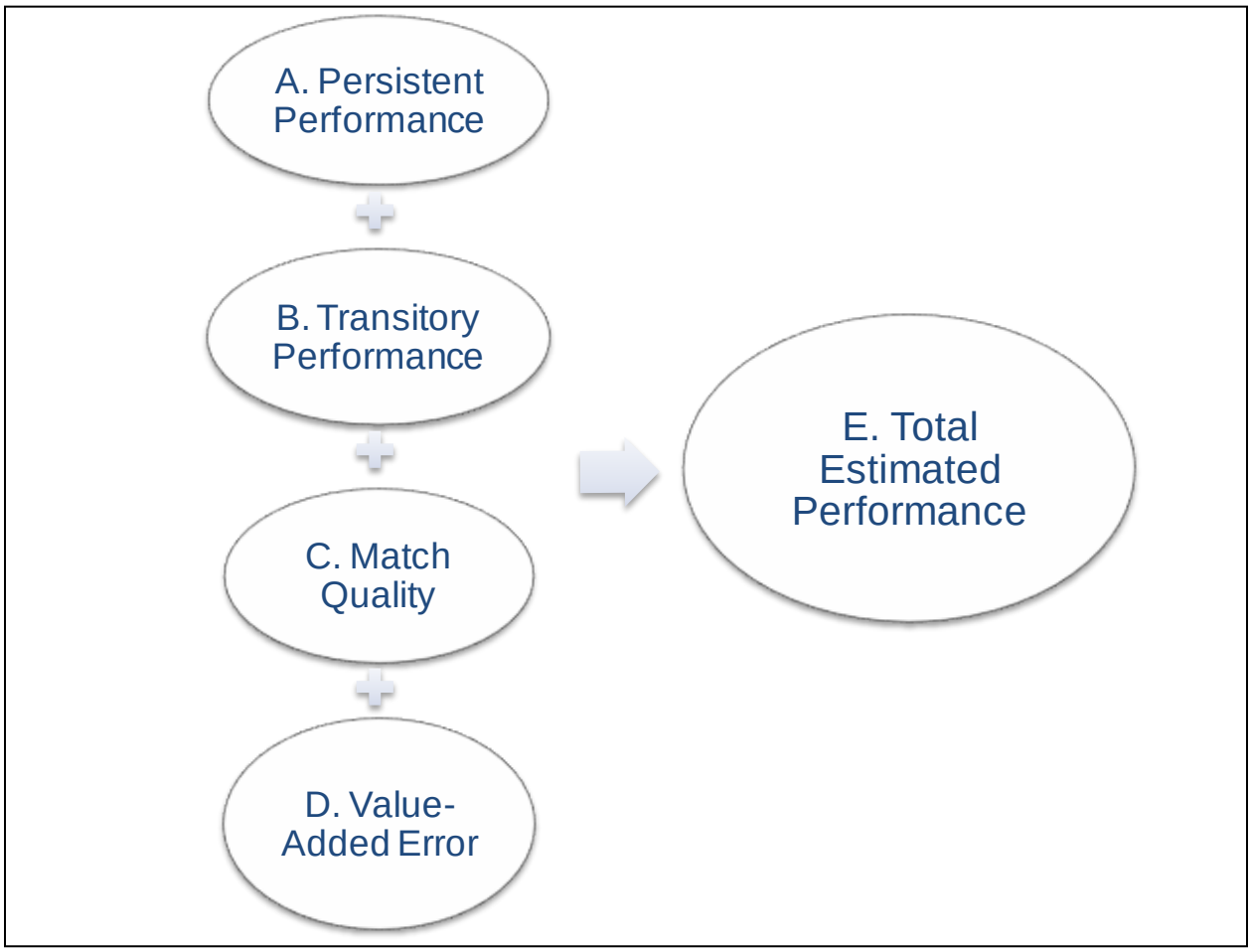
We hypothesize that estimated teacher performance is made up of four components, three of which represent aspects of true performance and one of which is measurement error. In Figure I.2, the first three components (labeled A, B, and C, respectively) are:

A. "Persistent performance" of a teacher, which can be thought of as teaching talent that is the same from year to year

B. "Transitory performance" of a teacher, representing whether the teacher was particularly effective or ineffective in a given year

C. "Match quality," which results in a positive effect if the teacher is matched with students that he or she is especially skilled at teaching or with other aspects of the position that align well with the teacher's skills, and a negative effect if the teacher is working with students or in an environment for which he or she is not as well equipped to teach

Figure I.2. Factors That Contribute to Estimated Teacher Performance



The final component, "value-added error" (labeled D), represents all the unmeasured factors that contribute to a teacher's value-added performance measure, such as luck and unmeasured influences of his or her students' family background on achievement growth. The greater the error relative to persistent performance, the greater is the probability of mis-classifying a teacher as highest-performing, and hence the weaker the expected effect of a transfer teacher.

If the persistent performance component is large relative to the others, we would expect to find an impact of the transfer incentive on student achievement, all other things being equal. If the student match component is large relative to the others, then the impact could be positive or negative, depending on whether the transfers resulted in an improvement or a weakening of the match between teacher talents and the types of students served by the schools to which those teachers transferred. If the transitory effects and random error are large relative to the other components, then the net effect would tend to drive the impacts down, toward zero.

The direction of the match effect is unknown. As noted above, TTI transfers must be completed by mutual consent. In order for the transfer to occur, transfer candidates must be willing to apply, interview, and accept a transfer position, while receiving-school principals must be willing to interview and offer a position to the transfer candidate. If teachers and principals consider the teacher-student match quality during the entire process of application, interview, offer, and acceptance of the offer, then match effects could be positive. However, there is a possibility that the difference in student characteristics would manifest as a negative match effect if teachers who appear effective with higher-achieving students are not well equipped to teach lower-achieving students.

It is important to recognize the possible existence of each of these components in order to interpret the study findings; however, policy interest is focused on the net effect of these forces in the context of a policy like a transfer incentive. Therefore, we focus on the overall impact of transfer teachers on their students in their new settings and do not attempt to isolate match effects. However, we will report on the types of students taught in the transfer teachers' original schools and in their new schools after they transfer.

Another possible weak link in the causal chain is the idea that targeted teams, those in low-achieving schools that start off with vacancies, cannot find good candidates to fill the vacant positions. That is, we do not always know the quality of the teacher who would fill the vacancy in the absence of a transfer incentive. It is assumed that low-achieving schools have difficulty finding strong teachers for these vacancies, but the methods by which vacancies are filled are not well known. This aspect of the problem, which we documented in a separate research brief (Glazerman and Max 2011), will be important to capture in the current study.

The logic model presented here has several implications for studying the transfer incentives. First, the model suggests that one should pay attention to factors such as the assignment of students to classrooms and the degree to which the amount of mentoring and other supports teachers received varies within grade level teams. Second, the model suggests that the intervention may have an impact at two levels of aggregation. The first is the team level, where a team is a group of teachers teaching the same grade and subjects in the same school. The team level captures all of the components that feed into the total effect: direct, indirect, and resource allocation effects. The second is the teacher level, specifically, the high-performing teacher who transferred into that teaching team.

The randomized study design, discussed later in this chapter, considers the ability to use an incentive to hire a teacher as the treatment, not the transfer itself. To capture the difference between, say, forced transfers and transfer incentives, we study all teaching teams where a vacancy was designated as eligible for the transfer incentive, regardless of whether it was actually filled by a TTI teacher. Thus the total impact of a transfer *incentive* is a weighted average of the impact of transfers plus the impact of teachers who filled vacancies outside the TTI mechanism, despite the presence of the incentive.

3. Where TTI was Implemented?

The TTI was first launched as a pilot in one school district in 2008 and was called Project RISE, an acronym for Reaching to Inspire Student Excellence. The pilot experience gave the implementation team an opportunity to observe responsiveness of teachers and principals to the communication materials and identify potential obstacles to implementation. Rebranded as TTI,

but with essentially the same design, the intervention was repeated in the pilot district with more schools and expanded to include six other school districts in five states in 2009.

These seven districts were selected on the basis of their ability and willingness to carry out the intervention. This included their ability to link student data to teachers over time and their desire to address what they felt were inequities in the distribution of highest-performing teachers, according to district leaders. The study sought out districts that were large—with at least 40 elementary schools—and economically diverse—with more than 10 elementary schools with a low percentage (fewer than 40 percent) of students eligible for free or reduced-price lunch (FRL) and more than 15 elementary schools with a high percentage (greater than 70 percent) FRL. The rationale for large, diverse districts is to study settings in which there is scope for addressing imbalances throughout the teacher labor market. For example, five of the seven districts are county districts and they include urban, suburban, and rural schools that serve students from both very high- and very low-poverty backgrounds with a wide range of achievement levels.

Three more districts were added in 2010 to enlarge the sample for the study. This report focuses on the first seven, referred to as cohort 1 districts; a future report will incorporate information from all 10 districts in cohorts 1 and 2.

C. Studying Teacher Transfer Incentives

1. Research Questions

The study addresses implementation and impact. This report focuses primarily on the implementation and intermediate impacts, the first two questions listed below. The third question listed below will be the focus of a future report.

- What was the TTI implementation experience with respect to the teacher recruitment process?
- What were the teacher placement results and intermediate impacts of TTI? For example, who filled the vacancies compared to those who would have filled the vacancies in the absence of the intervention? How did the intervention affect collaboration? How did it affect the allocation of resources within the school, such as assignment of students to teachers, teacher mentoring, and teacher leadership?
- What was the impact of TTI on teacher retention and student achievement?[5]

The methods for answering these questions include descriptive tabulations (for implementation questions) and causal analysis (for impact questions). The causal analysis relies on a random assignment procedure discussed next.

[5] This question is not addressed in the current report but will be addressed in a future report.

2. Study Design

To answer the impact questions we implemented a randomized controlled trial (RCT) in which we are able to compare outcomes within teacher teams that have the chance to fill a vacancy with a TTI teacher to outcomes for comparable teacher teams that filled a teaching vacancy through whatever route the school would normally pursue. Specifically, we grouped teacher teams that had vacancies in the same grade levels but were in different schools within the same district into blocks (typically pairs) that were matched on the basis of subject, and where possible, school characteristics. Then, we randomly assigned the teams within each block to a treatment or control group.

For the vacancies in teams assigned to the treatment group, the principals were offered the opportunity to interview and hire teachers that we previously identified as highest performing. For those assigned to the control group, the principals were asked to fill the vacancies as they normally would. This creates two groups that are, on average, identical in terms of student characteristics and school contexts. The only systematic difference between the two groups was that those in the treatment group were potentially taught by transfer teachers and the rest were taught by typical hires at the same types of schools. Comparing outcomes for these groups will generate unbiased estimates of the impact of TTI on student achievement.

To further strengthen the study design, we exploited the possibility that pairs of similar schools would have eligible teams at more than one grade level and the same grade levels in each school. In such schools we were able to assign teams in such a way as to ensure that both schools in the pair had one treatment and one control team. Control group contamination, sometimes called "spillover," is a possible concern with a study design such as this. Risk of contamination occurs when there is a teacher team assigned to the control group in the same school as another teacher team assigned to the treatment group. For example, the sixth-grade math teachers might be in the control group and sixth-grade language arts teachers, including a TTI transfer, might be in the treatment group. If the transfer teacher has a large effect on his or her students, it would be reflected in their math scores for the control team, artificially reducing the estimated impact on math achievement.

We sought to avoid contamination by imposing the following rule on the random assignment process. Treatment and control teams in the same school had to be separated by at least two grade levels in elementary schools, and at least one grade level and in a different subject in middle schools. As long as those teams were at least two grades apart (or one grade and a different subject, for middle school), then there was little danger of a transfer teacher influencing a control team in the same school. (52 out of 85 schools contributed just one teacher team to the study, so this risk was not present at most schools).

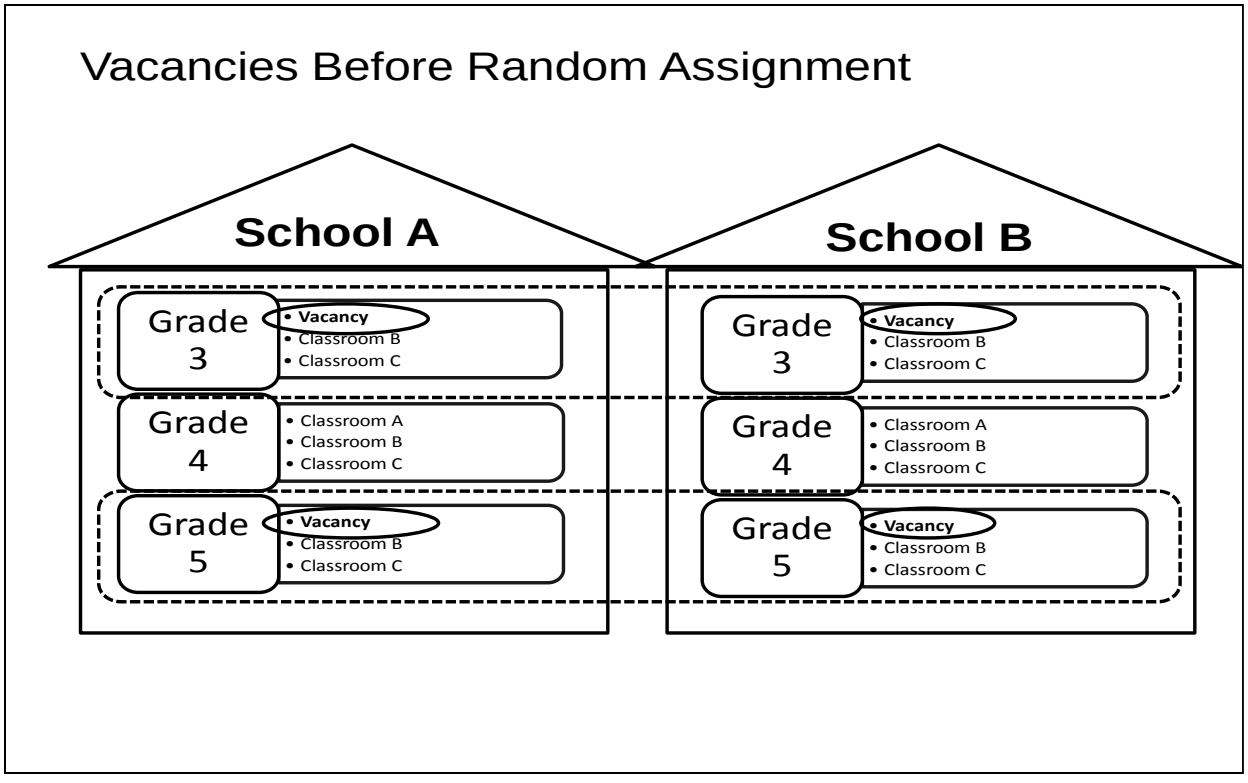
The random assignment design is illustrated for a hypothetical school pair in Figure I.3. In this example, two schools each have a teaching vacancy in grade three and another in grade five (top panel). In such a configuration, we assigned the third grade team in School A to either the treatment (TTI) or control group based on its random number and assigned the third-grade team in School B to the opposite status. Then we assigned the grade five teams in the respective schools to be the mirror image, so that each school had both a treatment team and a control team. The example in the bottom panel of Figure I.3 shows the result where grade three in School A and grade five in School B were assigned to have vacancies eligible for TTI, and grade three in School B and grade five in School A were assigned to the control group.

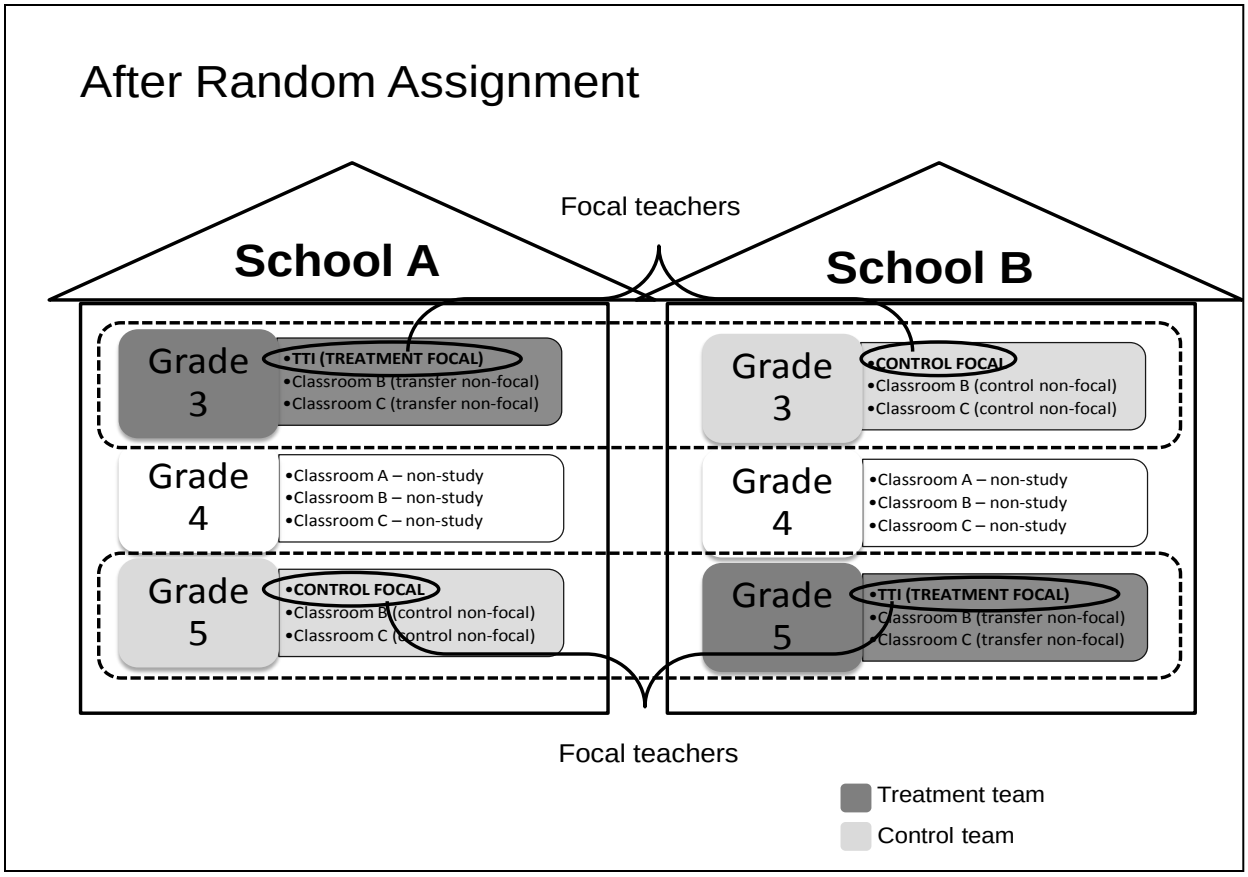
Another way to avoid contamination was to force teams into the same treatment status if they did not meet the adjacency rule described above. For example, if there were vacancies in grades three and four in an elementary school then both vacancies would be assigned to the same treatment status. If there were vacancies in sixth-grade math and eighth-grade math in a middle school then both would be assigned to the same status. Some teams had more than one vacancy in a single study team. In such cases (12 out of 124 teams), all vacancies within the team were assigned to a common study status because teams were the unit of random assignment.

As we will discuss in the future impact report, we obtain unbiased estimates of the total impact by comparing outcomes for TTI teams to outcomes for control teams. However, we expect much of the effect to operate through the teacher who filled the designated vacancy. We refer to this as the focal teacher. Therefore, we are interested in the team-level analysis as well as the comparison between focal treatment and focal control teachers.

Our study design included a careful determination of sample size in order to detect policy-relevant impacts. The size of the study sample was determined on the basis of being able to detect impacts on student test scores of approximately 14 percent of a standard deviation and being able to detect the effect of any factor affecting transfer probabilities by at least 5 percentage points. Because the current report includes only the first seven of the ten districts that ultimately participated in TTI, the results presented here should be interpreted with caution. Any finding that is not statistically significant may not necessarily mean that the program had no impact, but rather we were not able to detect an impact given the study's current sample size. This is particularly true for findings based on principal surveys. Principal survey questions can be broad in nature, and our aim for this report was to identify very large differences if they existed.

Figure I.3. Random Assignment Study Design





3. Data

We address the research questions through analysis of survey and administrative records data, as well as program implementation records. Surveys were conducted with teachers who were transfer candidates, regardless of whether they transferred; teachers in teams with vacancies, including both TTI ("treatment") teams and control teams of teachers; and their principals.[6] The administrative data include student test scores linked to teachers, demographic data, and teacher rosters.

Candidate Survey. In fall and winter 2009, the Candidate Survey was administered to the teachers eligible for the study.[7] The survey helped us characterize the background of teachers identified as highest performing and provided information about the factors affecting teachers' willingness to apply to the TTI, interview at low-achieving schools, and, ultimately, to transfer, as well as their experiences during the hiring process.

[6] The survey instruments are available at http://edicsweb.ed.gov/browse/downldatt.cfm?pkg_serial_num=4024.

[7] Teachers eligible for the study are the highest-performing teachers in each of the seven districts who are not already teaching in low-achieving schools or in schools that were exempted from the program by the district. Candidates who leave the district before being notified of the program opportunity are also excluded.

Teacher Background Survey. This survey was administered in late winter/early spring 2010 to all teachers in the study who filled one of the vacancies in treatment or control teams and to their colleagues in the same teaching team. It collected information on teachers' experiences at the study schools, along with information on their educational and professional background and other factors that may affect their students' achievement.

Principal Survey. The Principal Survey was administered in spring 2010 and spring 2011 to obtain data on teacher recruitment and hiring, principals' assessments of the teachers hired in the study's target grades, and any redistribution of resources across classrooms (including those related to the arrival of the new hire). The first-year survey collected information about hiring during the period of the TTI transfers. Emphasis in the follow-up survey was on teacher performance and school environment. (This report presents findings from the first round only.)

Teacher Roster Collection. Another outcome of interest is teacher retention. To compare retention rates of teachers in their new schools to those of existing teachers in the same schools, we collected teacher rosters for all study schools in the fall of each of the program's two school years and will do so again in the fall of the third year, after incentive payments are no longer being made. Retention findings will be presented in a future report on program impacts.

We obtained response rates of 83, 80, and 95 percent on the Candidate Survey, Teacher Background Survey, and Principal Survey, respectively. We received teacher roster data for 100 percent of the schools in the study. For the surveys, we conducted nonresponse analysis to describe the respondent samples and the degree to which each resembles the full population of respondents and nonrespondents.

For the candidate survey, the distribution of the respondents across districts and grade-subject pools (elementary, middle school English language arts, and middle school math) did not differ significantly from the distribution of the full sample of respondents and nonrespondents together. We also examined the percentages of respondent candidates who were placed in the top 10 percent of value added ranking in their grade-subject pool,[8] and found no significant difference with the percentages of candidates in the full sample (respondents and nonrespondents) who were in the top 10 percent. However, compared to the full sample of candidates, application and transfer rates among respondent candidates were significantly higher, which means that candidates who were interested in TTI are slightly over-represented in the candidate survey (Appendix A, Table A.1).

When we examined teacher and principal survey completion rates by district, grade, school poverty rate, school race/ethnicity, and school size there were few differences between the treatment and control group; 3 out of the 14 district comparisons and 1 of the 10 school race/ethnicity comparisons were significant (Appendix A, Table A.2). We also compared the respondents to the full sample of respondents and nonrespondents and found that distributions of sample members across each of the same groups (district, grade, school poverty, etc.) did not differ significantly between the subsample of respondents and the full sample except for grade and school race/ethnicity for the teacher survey (Appendix A, Table A.3). In those cases, despite

[8] All transfer candidates are in the top 20 percent of the value-added ranking in their grade-subject pool.

the statistical significant of the differences, the respondents had less than two percentage point differences in each category of grade level and less than two percentage point differences in each category of school race/ethnicity. To account for the possibility of nonresponse bias resulting from an over-representation or under-representation of certain groups, we controlled for group characteristics in all impact analysis results, including the experimental impact analysis.

D. This Report: A Focus on Implementation and Intermediate Impacts

The remainder of this report describes the implementation of TTI in the seven school districts that began the intervention in 2009 and finished in 2011. We focus on the implementation questions listed above, and also describe the study sample which will contribute to an impact analysis in a future report that combines these seven districts with three additional districts that began implementation in 2010. By focusing on implementation, this report will provide important information for policymakers and district administrators wishing to understand issues associated with teacher transfer incentives.

Chapter II describes the formation of the study sample. Chapter III documents implementation information from the period of intensive activity that takes place during the spring and summer leading up to the transfers, when transfer teacher candidates are identified and recruited, and principals at the receiving schools interview and make offers to them. Chapter IV describes the teachers who transferred, the teachers who would fill the vacancies in high-need schools in the absence of a transfer incentive, the degree of collaboration and interaction among teachers, and the changes that happen inside the receiving school as a result of the transfer incentive. The last two phenomena—collaboration and changes in resource allocation—relate to the indirect and resource allocation effects discussed in the logic model above. Finally, Chapter V provides a brief summary of findings and next steps for the study.

II. THE STUDY SAMPLE

This chapter describes the study sample, including how the school districts and schools were selected and recruited into the study. We also describe the policy environment in which the study operated. To follow the discussion, it helps to know that we partitioned elementary and middle schools in each participating district into three groups:

- **Potential Receiving Schools**. These are the lowest-achieving schools in the study districts, into which teachers can potentially transfer.
- **Potential Sending Schools**. These are higher-achieving schools, from which teachers can potentially transfer.
- **Exempt Schools**. Approximately nine percent of schools across the seven study districts were exempted because they were already participating in a comparable intervention or primarily serving special populations.

A. Selection, Recruitment, and Description of School Districts

1. District Selection

We selected school districts that were large and economically diverse. Large districts are necessary because the intervention is more likely to be feasible with a large pool of sending and receiving schools to consider. The study required enough schools not only to implement the intervention but to form a control group. To be "large," districts had to have at least 40 elementary schools. We identified 59 districts that met this initial screening criterion. Economic diversity is also important, because the Talent Transfer Initiative (TTI) encouraged the transfer of teachers from high-achieving to low-achieving schools and we hypothesized that these gaps would be starkest when there were large income disparities throughout the school system. We used the number of high- and low-poverty schools, using free or reduced-price lunch (FRL) eligibility as a proxy, to determine whether a district had a sufficient mix of schools at different achievement levels.[9] Districts were deemed to be economically diverse if they had at least 10 low-poverty elementary schools (that is, schools with less than 40 percent of students eligible for FRL) and at least 15 high-poverty elementary schools (that is, schools with greater than 70 percent of students eligible for FRL). Out of the 59 districts that met the size criterion, 51 districts were deemed to be economically diverse based on these requirements.

In addition to the quantitative criteria, we created a prioritized list of these 51 districts based on a variety of factors, including test score availability, data quality, hiring/transfer practices, and the local political environment. These factors affected the feasibility of conducting the program in a district. Information used in this process was based on data gathered by the three organizations represented on the recruitment team: Mathematica, The New Teacher Project (TNTP), and Optimal Solutions Group.

[9] In six of the seven study districts, the correlation between school-level achievement rank and FRL rates ranged from -0.70 to -0.94. The correlation in the seventh district was -0.40.

Of the 51 prioritized district candidates, we recruited seven districts for cohort 1 (beginning in the 2009–2010 school year) and another three districts for cohort 2 (beginning in the 2010-2011 school year). We selected these particular districts by starting with the largest districts and attempting to contact them in order, excluding any that were unwilling to participate. We prioritized 19 districts to achieve the sample of 10.[10]

2. Participating Districts and Study Context

In order to interpret the findings of this study, one must understand the context in which it operated. This includes the overall characteristics of the specific districts participating, the economic and geographic conditions facing teachers in those districts, and the compensation policies affecting such teachers.

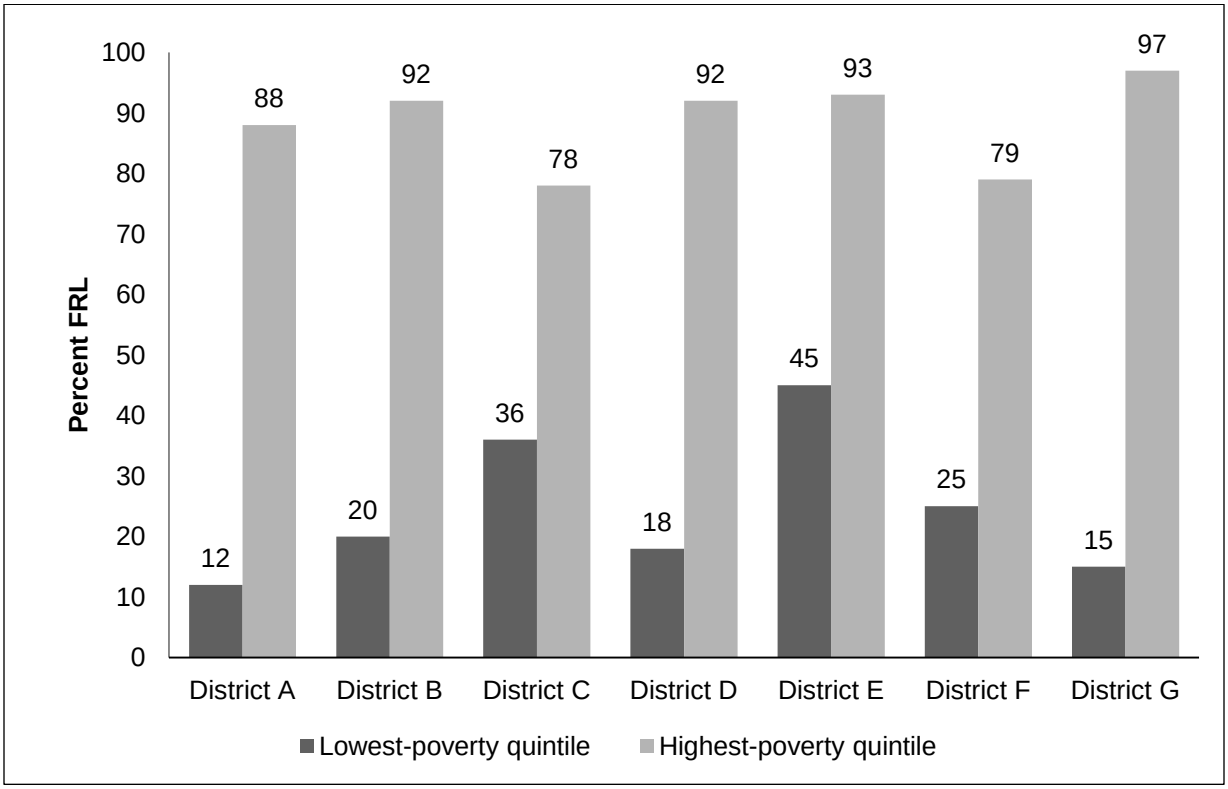
a. Characteristics of Participating Districts

Although all of the study districts were large (with at least 40 elementary schools), they varied in size and student characteristics. The largest of the study districts had more than three times as many elementary schools and four times as many elementary school students as the smallest. The proportion of students in districts who were African American or Hispanic ranged from 19 percent to 89 percent. Across the seven districts, between 46 percent and 71 percent of students in elementary schools were FRL eligible.

It is also important to consider the variability in FRL rates among schools, because economic diversity of schools was a district selection criterion. Within each district, we ranked all of the schools by the percentage of students FRL eligible, and divided the list into five equal-sized groups (quintiles) separately for elementary and middle schools. Figure II.1 shows the spread of low-income students between the lowest and highest poverty quintile of elementary schools. For example, in District A, there was a 76 percentage point difference in the FRL rate between the average schools in the top and bottom quintile of elementary schools. Other districts had gaps that ranged between 42 and 82 percentage points. Figure II.2 shows the corresponding results for middle schools.

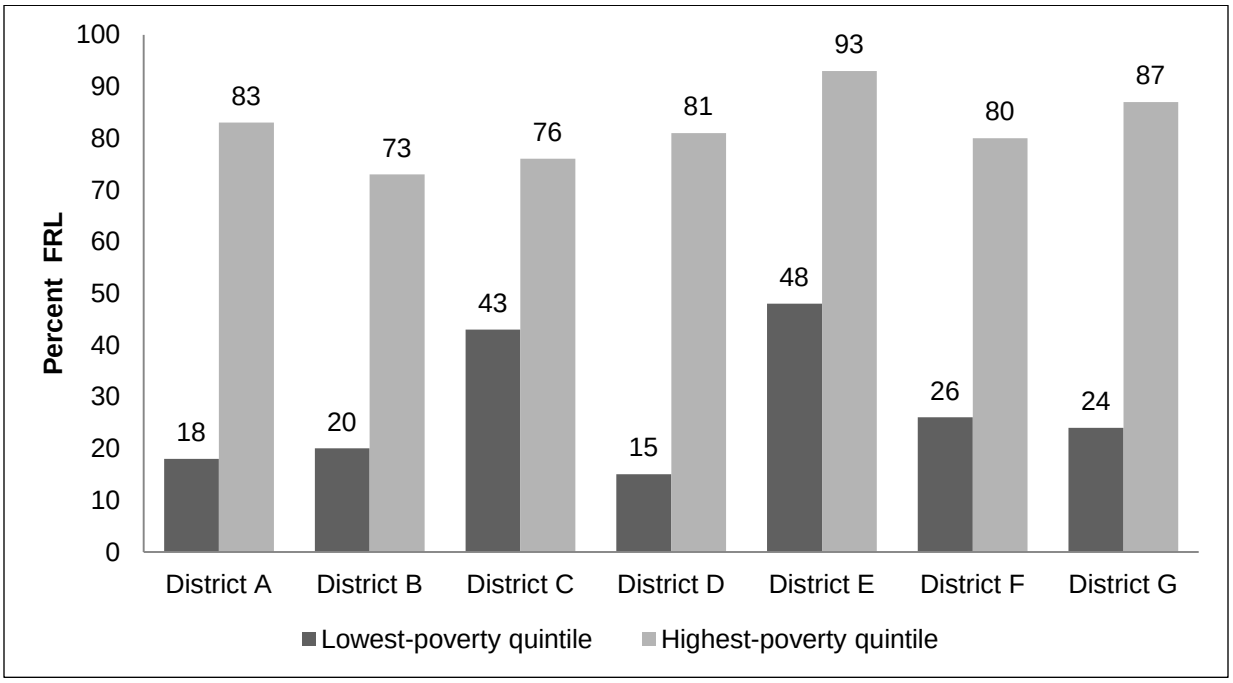
[10] Because the districts had to volunteer to participate, they should not be considered a statistically representative sample of those identified by the initial, quantitative selection criteria.

Figure II.1. Percentage of Low-Income (FRL) Students in Lowest- and Highest-Poverty Elementary Schools



Source: National Center for Education Statistics Common Core of Data (http://nces.ed.gov/ccd/).

Figure II.2. Percentage of Low-Income (FRL) Students in Lowest- and Highest-Poverty Middle Schools



Source: National Center for Education Statistics Common Core of Data (http://nces.ed.gov/ccd/).

b. Employment Landscape and Geography of Participating Districts

Labor market conditions for teachers as well as each district's physical geography are important factors that could affect the implementation of a transfer incentive intervention. To set the context for the current study, we examined these conditions in the seven participating districts. All of the seven TTI districts are located in states with "right-to-work" laws, where teachers must each decide affirmatively whether to join or pay dues to a union.

While union strength is one aspect of the teacher labor market, another is the general unemployment rate. During recruitment for the pilot study (2008), the national unemployment rate was 5.8 percent, but by 2009, the year when cohort 1 districts began full-scale implementation, the rate had risen to 9.3 percent. The local unemployment rates for each district followed a similar pattern.[11]

The geographic features of the districts are also relevant to understanding teachers' willingness to transfer. The largest of the TTI cohort 1 districts is 1,233 square miles, which is larger than the state of Rhode Island (1,045 square miles). The other six districts range in size from 195 to 650 square miles. Respondents to the Teacher Background Survey reported that their average commute to school was 13.4 miles each way for an average of 21.3 minutes.

c. Existing Incentive Programs in Participating Districts

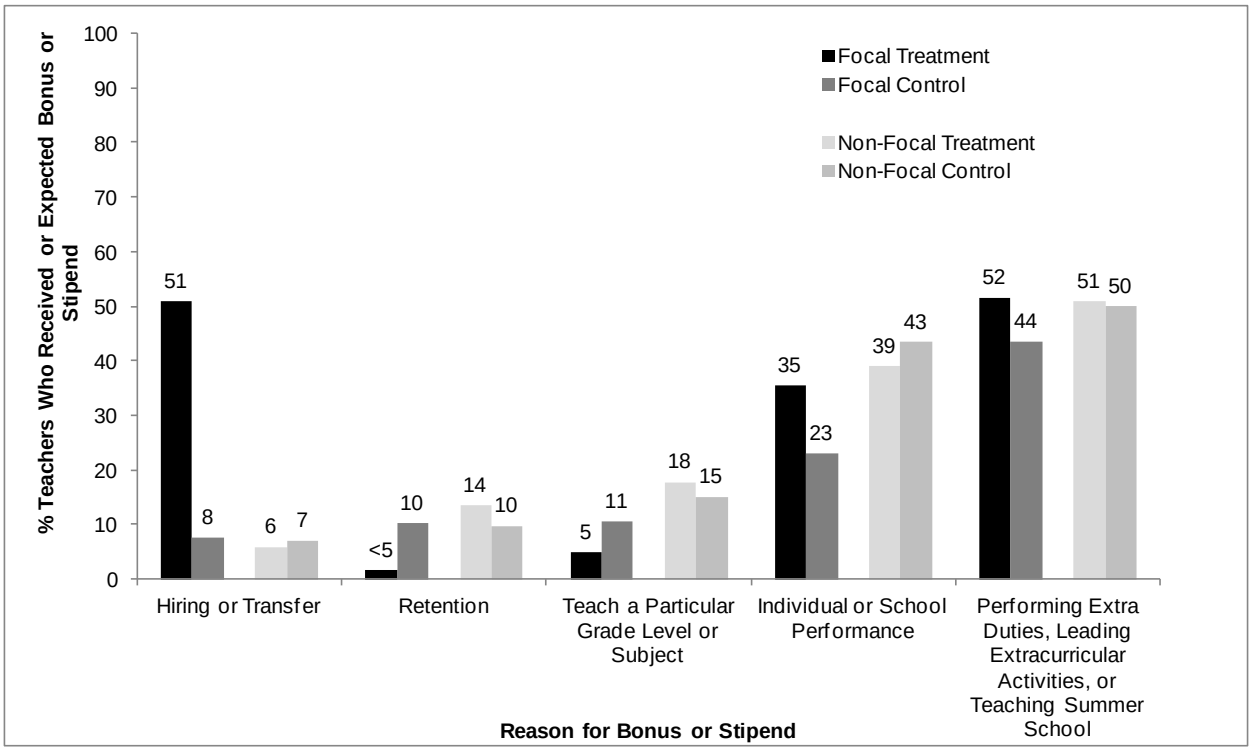
Although we excluded any school districts where existing or planned teacher incentive programs would have duplicated the intervention under study, we did encounter some existing policy initiatives in each of the seven participating school districts. These programs included performance incentives and signing bonuses for teachers. In each case, we determined that the existing programs were different enough, isolated to a few schools that could be excluded from the study of TTI, or involved small enough dollar amounts that they would not interfere with the study design. These incentive programs were funded by a variety of federal, state, and district sources. Teachers and schools receiving more than $5,000, an arbitrary threshold used to identify substantial bonus programs, were excluded from the study in order to reduce complicating the study by changing the effective incentive offered by the TTI intervention and the counterfactual. Only one district had an intervention that was very similar to TTI, and that intervention only affected a small number of teachers and schools, which were excluded from the study. Since any competing program has the potential to shrink the pool of transfer candidates, we will interpret the results of this study in light of the existence of these programs in the final report.

[11] Based on data from the U.S. Bureau of Labor Statistics (http://bls.gov/lau/#tables), we found that the unemployment rates in TTI study districts ranged from 4.8 to 6.7 in 2008 and rose in every district in 2009, ranging from 7.2 to 10.8.

d. Bonuses, Stipends, and Additional Payments Received by Teachers

To interpret the findings of the study, it is important to understand how the $20,000 incentive ($10,000 per year) offered to teachers filling vacancies in the treatment group relates to any other payments being offered to teachers for hiring, retention, or performance. As part of the Teacher Background Survey, we asked teachers in study grades about the bonuses and incentives, including any offered by TTI, that they were offered during the first year of the study.[12] To better understand the policy environment in which the TTI operated, we summarized in Figure II.3 the responses of teachers in study grades. There are four groups of teachers in study grades presented in this figure: focal treatment teachers, most of whom received TTI transfer stipends; focal control peers, who filled control study positions and were not eligible for TTI stipends; and non-focal treatment and control peers, who taught on the same teams as focal teachers but did not fill study positions and also did not receive TTI stipends.

Figure II.3. Percentage of Teachers Offered Bonuses and Stipends in 2009-2010

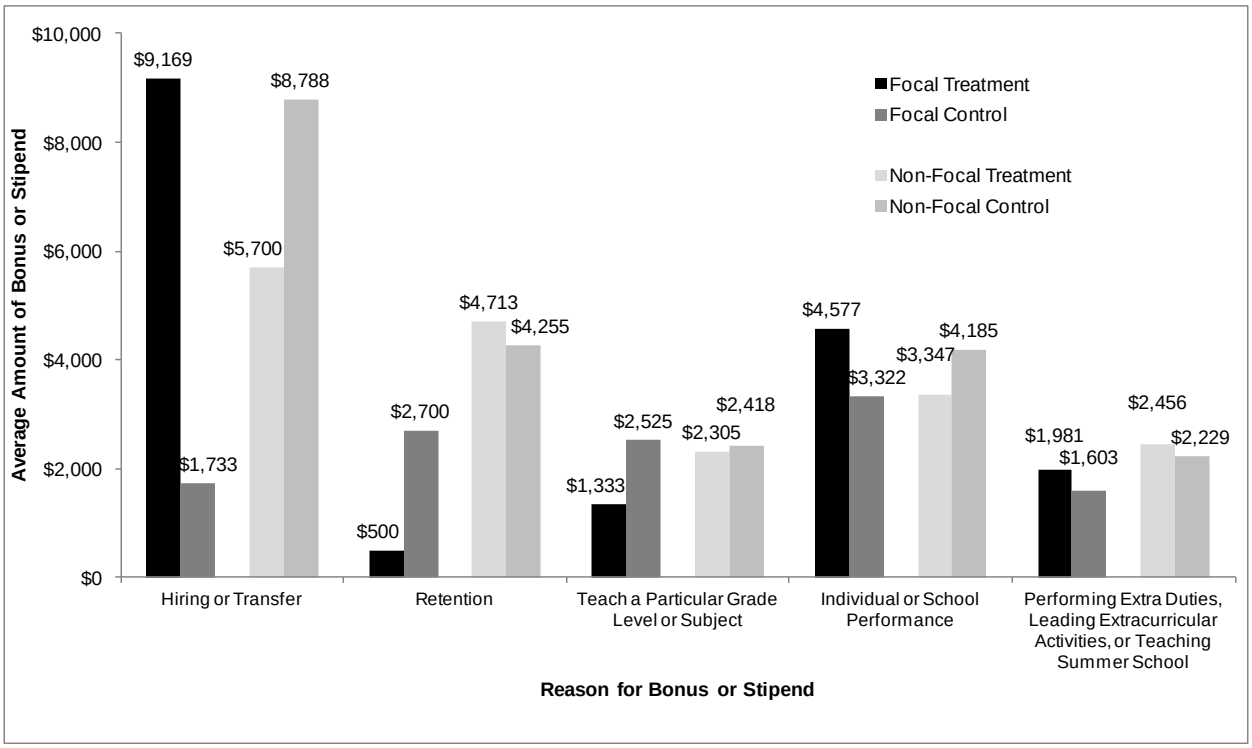


Source: 2009–2010 Mathematica Teacher Background Survey.

[12] The question asked teachers if they had received or had been offered various types of bonuses and stipends. Some teachers may have answered affirmatively that they were offered a bonus but did not actually receive it.

Bonuses similar to those offered by the TTI were relatively rare. With the exception of treatment focal teachers, fewer than 10 percent of respondents in each category reported being offered a hiring or transfer bonus. Across all categories, fewer than 20 percent of respondents reported being offered a retention bonus or a bonus to teach a particular grade level or subject.[13] The average amounts of the bonuses are summarized in Figure II.4. On average, bonus amounts were less than $5,000 for all types of bonuses except for hiring and transfer bonuses. The average hiring and transfer bonus amounts for non-focal teachers are greater than $5,000, but this average is inflated by four teachers in two districts who reported $20,000 bonuses. The majority of non-focal treatment and non-focal control teachers received less than $5,000 in hiring and transfer bonuses.

Figure II.4. Bonuses and Stipend Amounts Offered to Teachers in 2009-2010



Source: 2009–2010 Mathematica Teacher Background Survey.

It should be noted that although 85 percent of treatment focal teachers received TTI transfer stipends of $10,000 in 2009-2010, only 51 percent reported having been offered transfer bonuses in this year. It is possible that some of these teachers considered their TTI stipends to be performance-based bonuses rather than transfer bonuses, as their past performance was emphasized in the recruiting stage. Indeed, 29 percent of transfer teachers who reported not receiving transfer stipends did report receiving performance-based bonuses. Another possibility is that they thought that their TTI stipend was offered in 2008-2009, because they were technically offered the stipend in that year even though they were paid in 2009-2010.

[13] These numbers include any teachers who were receiving $5,000 per year in retention stipends through the TTI because they were deemed highest-performing and were already teaching in low-achieving schools.

B. Selection of Sending and Receiving Schools

We partitioned the set of elementary and middle schools in each district into three groups: potential receiving schools, potential sending schools, or exempt schools. This was necessary because TTI was designed so that only the lowest-achieving schools (receiving schools) were offered the opportunity to hire high-performing teachers through TTI. With the exception of schools that we exempted from TTI because of special circumstances, the remaining schools in the district formed the pool of potential sending schools.

After describing the *potential* sending and receiving schools below, we go on to describe the *participating* sending and receiving schools, in other words, the schools that transfer candidates actually left and the schools to which they went.

1. Identifying Potential Sending and Receiving Schools

Before identifying sending and receiving schools, the districts first listed schools that would be exempt from being either a sending or a receiving school because they were already receiving a comparable intervention[14] or primarily serving special populations.[15] In addition, one district chose to exclude low-achieving schools that had shown strong learning gains. Overall, districts participating in TTI excluded nine percent of all schools from the program.

After exempting these schools, the primary criterion for potential receiving schools was that they must be low achieving based on average student test scores in the grades and subjects targeted by the program (math and reading in grades three through eight). Working closely with each district, Mathematica ranked schools in each district, separately for the elementary and middle school levels, on the basis of average student achievement in reading and math using either the prior three years or the prior one year of achievement data, depending on the district leaders' preferences.[16] To ensure consistency with the school accountability system in each district, student achievement data came from the annual assessments used by these systems.

Two districts chose to incorporate additional information from their school accountability systems into the school selection process. One district sorted schools by accountability rating first and then by average student achievement, allowing it to focus on schools with the two lowest accountability ratings. The ranking based on average student achievement differed slightly from accountability ratings because the accountability ratings included information on average achievement for student subgroups and on achievement in social studies and science. As

[14] Adding the TTI bonus to existing bonuses would not only risk duplicating services, but would also complicate the study. It would alter the treatment and control conditions, making the treatment a larger incentive amount than in other districts and the counterfactual a different recruitment incentive, rather than no incentive. Consequently, schools offering existing bonuses of $5,000 or greater were excluded from the sending and receiving pools.

[15] Special populations include students who are blind or deaf or students with severe learning disabilities. These schools often required teachers to have special training or certification and were not appropriate for TTI transfers.

[16] Achievement data from the year prior to the implementation of TTI was used for all but two districts, where 3 prior three years of achievement data were used.

a result, some schools in this district were classified as potential sending schools based on their accountability rating even though they had lower average achievement than some potential receiving schools. Another district excluded schools already receiving resources through a program that targeted schools with low accountability ratings.

The cutoff between sending and receiving schools had to be carefully set in order to obtain both a sufficient number of vacancies in receiving schools and an adequate pool of eligible highest-performing teachers for transfer from sending schools. To achieve this tradeoff, the project team drew upon the experience from the pilot study to obtain a sufficient number of vacancies and an adequate pool of eligible highest-performing teachers for transfer.

Overall, 21 percent of schools across all seven districts were identified as potential receiving schools, with 70 percent potential sending schools, and the remaining 9 percent exempt. Thus, the ratio was 3.3 potential sending schools per receiving school, with a range across districts of 2.4 to 4.0. In terms of numbers of schools, we identified 154 potential receiving schools, an average of 22 per district (ranging from 13 to 42 schools, by district).

Consistent with the program design, potential receiving schools were more disadvantaged than potential sending schools. We used the percentage of students eligible for FRL as a measure of disadvantage. In elementary schools, 78 percent of students in the average receiving schools were FRL compared to 64 percent of students in sending schools, a statistically significant difference of 14 percentage points.[17] For middle schools, the difference was also statistically significant, equal to 20 percentage points (74 versus 54 percent).

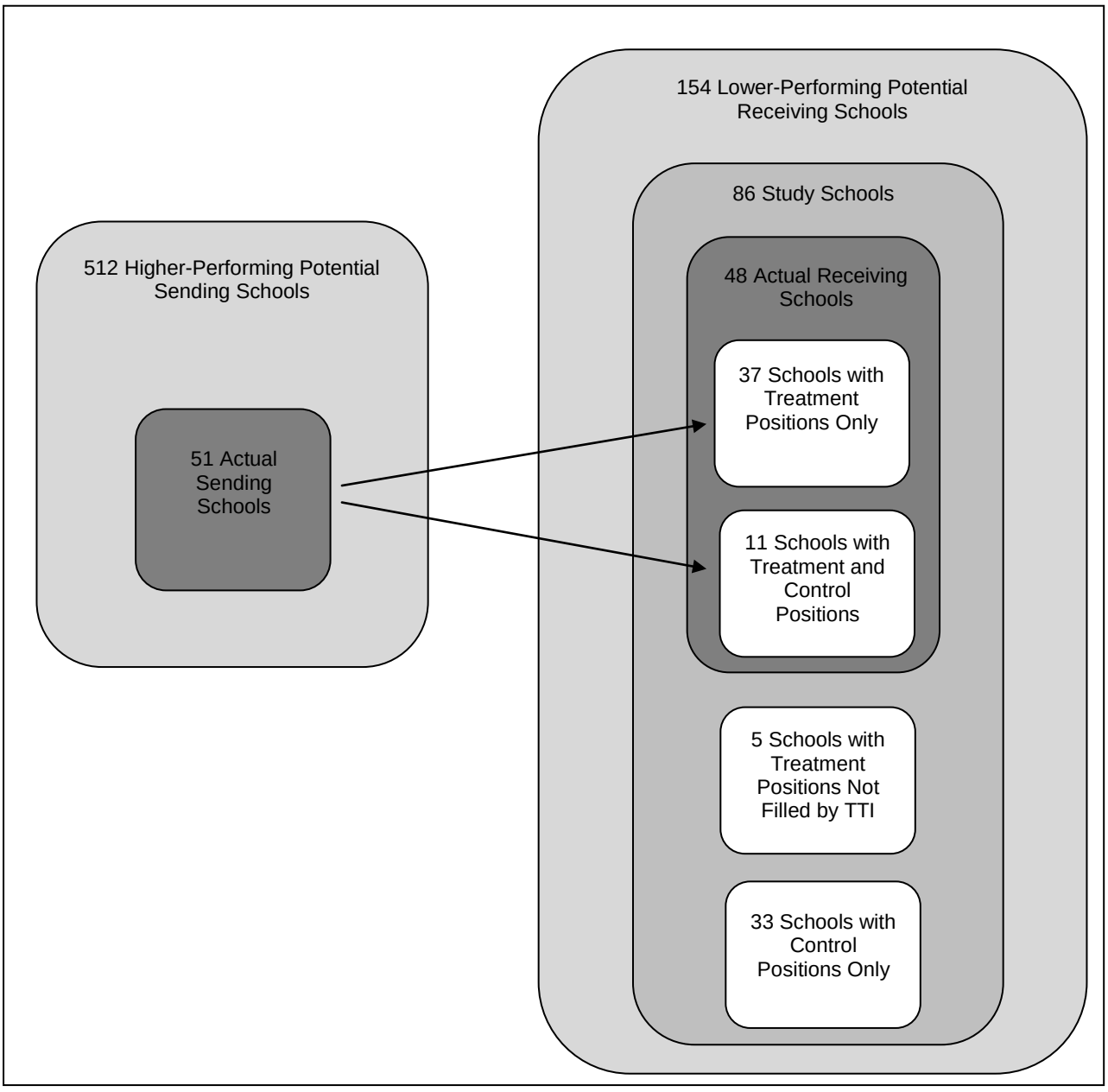
2. Participating Sending and Receiving Schools

Not every potential sending school had a teacher transfer out. Figure II.5 shows that, across the seven districts, TTI transfer candidates came from 51 out of 512 (10 percent) of the potential sending schools. This percentage is relevant for districts concerned that a transfer program might be disruptive to many of its sending schools. In fact, 90 percent of schools identified as potential sending schools did not lose any teachers through the transfer program.

Not every potential receiving school had a teacher transfer in. Figure II.5 shows that the TTI teachers transferred into 48 potential receiving schools. These 48 schools represent 91 percent of the 53 schools with a vacancy that had been assigned to the treatment group. The 5 schools that had vacancies assigned to treatment group but did not receive any transfer teachers chose to fill their vacancies outside of the TTI pool or did not fill the position with any teacher due to cuts. The other 101 potential receiving schools included both those with no vacancies submitted for random assignment (68 schools) and those where all of their vacancies submitted to the study had been assigned to the control group (33 schools).

[17] Discussion of "significant differences" here and throughout refers to statistical significance. We use a 0.05 significance level, which means that a significant difference is highly unlikely (less than five percent of the time) to be observed in a sample if the population difference was zero. Statistical significance does not imply that the difference is meaningful to policy, nor does a lack of statistical significance imply that the difference is not meaningful for policy.

Figure II.5. Sending and Receiving Schools



The study focuses on selected grade-subject teams within participating receiving schools.[18] Participating receiving schools are those where the team had been assigned to the treatment or control group. This study sample has teams from 86 schools, including the 48 with TTI transfers, 5 with teams assigned to TTI but had no transfers, and 33 with only control teams. Treatment vacancies that were not filled through TTI are considered to be in the treatment group in this study, following an intent-to-treat approach. Next, we describe the study sample in terms of the students in those grade-subject teams and the teachers who had already been teaching on these teams when the study began.

C. Treatment and Control Teams: Baseline Characteristics of Students and Teachers

Baseline differences in student background between the treatment and control groups were not statistically significant. Table II.1 shows the pre-test scores and demographic characteristics of students on treatment and control teams. Both groups had scores from their prior year that were below the state average by at least one-third of a standard deviation, placing the average sample member below the 37th percentile in both math and reading.

We also examined the demographic and professional characteristics of teachers on treatment and control teams at baseline (see Table II.2). These data were collected from teachers through the Teacher Background Survey. In order to capture baseline characteristics, the teachers included in the analysis in Table II.2 are the non-focal teachers, in other words, the peers of the teachers who filled vacancies on treatment and control teams. The TTI teachers on treatment teams and the focal control teachers are not included in this analysis, since TTI has the potential to affect differences in focal teacher characteristics.

Table II.1. Team-Level Mean Student Characteristics, by Treatment Status

Student Characteristic	Treatment Mean	Control Mean	Difference	P-value
Prior Achievement[a]				
Math pretest score	-0.36	-0.37	0.01	0.918
Reading pretest score	-0.42	-0.36	-0.06	0.383
Demographics (percentages)				
Male	52.1	51.9	0.2	0.903
Race/ethnicity				
White	8.4	6.4	1.9	0.322
African American	35.5	47.9	-12.4	0.051
English language learner	35.7	28.4	7.3	0.155
Special education	16.3	20.6	-4.3	0.119
Free or reduced-price lunch	77.5	71.9	5.7	0.129
Sample Size (students)	**8,039**	**7,127**		

Source: Administrative data.

Note: None of the differences is statistically significant at the 0.05 level using a two-sided test.

[a]Test scores are reported in standard deviations relative to the state average.

[18] The research team decided that, with limited evaluation resources, collecting data on sending school teacher transitions would be lower priority than collecting data on outcomes for the receiving schools.

There were no statistically significant differences in terms of professional qualifications between non-focal teachers on treatment and control teams. The only statistically significant difference in terms of demographic characteristics was that, on average, non-focal teachers on treatment teams were less likely to be married or have children than non-focal teachers on control teams.

Table II.2. Team-Level Mean Non-Focal Teacher Characteristics, by Treatment Status

Teacher Characteristic	Treatment Mean	Control Mean	Difference	P-value
Personal Characteristics				
Female (%)	81.6	83.0	-1.5	0.775
White (%)	39.5	34.8	4.7	0.472
African American (%)	37.7	39.3	-1.6	0.810
Hispanic (%)	19.3	24.1	-4.8	0.383
Age in 2010	40.5	41.6	-1.1	0.494
Married or with partner (%)	44.7	64.9	-20.1*	0.002
Have children (%)	40.4	56.4	-16.0*	0.016
Homeowner (%)	69.3	73.2	-3.9	0.518
Professional Characteristics				
Years of teaching	9.9	11.3	-1.4	0.198
Years teaching in district	8.0	9.2	-1.2	0.187
Years teaching at school	5.5	6.3	-0.8	0.260
Master's degree or higher (%)	43.9	41.1	2.8	0.673
National Board Certified (%)	16.4	21.5	-5.1	0.337
Grade certified (%)	93.9	89.1	4.8	0.202
Sample Size (teachers)	**114**	**114**		

Source: 2009–2010 Mathematica Teacher Background Survey.

*Difference is statistically significant at the 0.05 level using a two-sided test.

This page has been left blank for double-sided copying.

III. THE TRANSFER PROCESS

In this chapter, we describe in detail the process of implementing the Talent Transfer Initiative (TTI) through all steps of the transfer process, from identifying the highest-performing teachers through their placement and acceptance of positions in low-achieving schools. To receive the $20,000 transfer incentive, a teacher had to be identified as one of the highest-performing teachers in the district, be teaching in one of the designated potential sending schools, apply to the Talent Transfer Initiative (TTI), interview, be offered and accept a position from a receiving-school principal, and finally, remain in his or her school and grade-subject team for the next two years. Because this process was both voluntary and competitive, the outcomes depended on the behavior of both the teachers identified as highest performing (transfer candidates) and the principals in the low-achieving (potential receiving) schools.

A. How Were Transfer Candidates Identified and Recruited?

1. Value-Added Analysis to Identify the Highest-Performing Teachers

The first step in identifying transfer candidates was to use value-added analysis to identify each district's highest-performing teachers. As discussed in Chapter I, value-added analysis uses student achievement growth data to identify the contribution of each individual teacher to his or her students' growth, holding constant as many factors as possible that are outside the teacher's control. We considered three separate pools of teachers in each district—elementary school multiple-subject, middle school English language arts (ELA), and middle school math teachers. In four of the seven districts, Mathematica conducted the analysis (see Appendix B for technical details), which took place between January and March 2009. The three other districts were already using value-added systems developed by an external partner, the SAS Institute, to evaluate teacher performance. For two of those three districts, we used value-added scores provided by the districts from their external partners to identify the highest-performing teachers. The remaining district provided us with a list of teachers (rather than scores) that the district itself had identified as highest performing based on value-added scores provided by their vendor.[19]

The value-added estimates were based on two to three years of test data and control for students' prior achievement and background characteristics (See Appendix B). Using more than one year of data allows for a better estimate of teachers' "persistent performance" and reduces the influence of transitory performance, mentioned in the logic model in Figure I.1 in Chapter I. However, it also reduces the number of teachers who can qualify as high performing. Requiring two or more years of data meant that teachers who were new to the district were not eligible. Teachers who taught in only one of the three years prior to the implementation of TTI or left

[19] The value added analyses conducted by the SAS Institute and by Mathematica were not coordinated with each other in any way. The goal was to use value added models that would plausibly be used by a district seeking to implement an intervention such as TTI.

before the current academic year began were also ineligible.[20] For the four districts where we estimated value-added scores, we used three years of student test scores, demographic characteristics, and enrollment data from school years 2005–2006 to 2007–2008. Individual teachers who had taught two of the three years could still qualify. In the two districts that provided value-added scores, one district provided scores for those same three years and the other provided scores for two years.

Among the eligible teachers, those whose value-added scores placed them in the top 20 percent in their district and pool—elementary school multiple-subject, middle school English language arts (ELA), and middle school math teachers—were identified as highest performing and were designated as transfer candidates for the purposes of the TTI. As discussed in Chapter II, this cutoff was chosen so as to be selective yet provide enough transfer candidates to yield adequate numbers of program applicants to fill all of the vacancies identified for receiving schools. Based on experiences from the pilot study, the goal was to identify at least 10 candidates per vacancy to be filled with a TTI teacher. The choice of 20 percent as the arbitrary cutoff usually generated a large enough pool to fill the target number of vacancies. (As shown later, 90 percent of the vacancies were filled). We adjusted the cutoff for some pools in some of the districts either to be more selective or to slightly enlarge the pool of candidates. Specifically, it was lowered to 18 percent for elementary teachers in two districts, raised to 23 percent for middle school math teachers in one district, and raised to 25 percent for middle school teachers in one other district.

a. Who Was Identified as Highest Performing?

Across the seven districts, 1,012 teachers were identified as candidates for the 63 positions that were ultimately filled, a ratio of more than 16 candidates per position.[21] Table III.1 compares value-added scores of the highest-performing teachers to the other eligible teachers in the four districts in which we estimated value-added scores. Our value-added model estimated teachers' contribution to student achievement growth—value added—in terms of standardized student test scores, scaled so that a score of 1.0 represents one standard deviation above the mean for the distribution of test-takers (students) district-wide in each respective district. By construction, this scaling results in a value added of zero for the average teacher in the analysis sample for a given pool within a district. Also, the value added of any given teacher is the amount of extra progress (if positive) that the teacher's students made with him or her relative to the average teacher in terms of district-level student standard deviation units.

[20] The percentage of teachers eligible to be considered for high performing ranged from 35 to 65 for elementary school teachers across districts, and from 22 to 70 for middle school teachers. In other words, there was a pool of teachers where 78 percent of teachers that we identified who had ever taught a student in that pool did not teach a sufficient number of students in the same pool consistently for three years in a row.

[21] We initially identified 1,385 teachers as highest performing, but some were no longer teaching or not planning to teach in the year during which the program sought to have them transfer. Counting the teachers who turned out to be ineligible for the TTI, the ratio of candidates to filled vacancies was almost 22. These numbers are approximate because we only had sufficient data to count initially identified teachers for six of the seven districts. We used the ratio for the six districts (1,286 total to 940 eligible) and multiplied it by 1,012, the number of eligible teachers in all seven districts, to extrapolate the estimated total for all seven districts.

By definition, average value-added scores for the highest-performing teachers are higher than those of the other eligible teachers. The mean value added for the highest-performing teachers of all grades together (grades 3–8) for ELA was 0.13 standard deviations above the value added of the average teacher; for math it was 0.24 standard deviations above. The mean value added for all other eligible teachers was significantly lower: 0.16 standard deviations below the value added of the average highest-performing teacher for ELA, and 0.28 standard deviations below the average highest-performing teacher for math. Translating standard deviations to percentiles, the average highest-performing teacher would move his or her students up by an average of 6 percentile points for ELA and 11 percentile points for math in a school year compared to the average teacher among those not highest performing in the district.[22] Table III.1 presents the mean value added for these groups separately as well as by grade span. As mentioned, all of these results are based on four of the seven districts that provided us with detailed data. The magnitude of these differences in average value-added scores may not necessarily be similar for the other three districts.

Table III.1. Value-Added Scores: Highest-Performing vs. Other Eligible Teachers

	Highest-Performing Teachers (Top 20%)†		Other Teachers (Bottom 80%)†		
	Mean	Sample Size	Mean	Sample Size	Difference
All Grades					
English language arts	0.13	621	-0.03	2,412	0.16*
Math	0.24	599	-0.04	2,294	0.28*
Elementary (Grades 3-5)					
English language arts	0.13	448	-0.03	1,727	0.16*
Math	0.24	448	-0.04	1,734	0.28*
Middle School (Grades 6-8)					
English language arts	0.12	173	-0.02	685	0.14*
Math	0.22	151	-0.03	560	0.25*

Source: Estimation by study team from administrative data.

Notes: Data pertain to a subgroup consisting of the four districts whose value-added estimates were calculated by the study team.

† Value-added scores are in student-level standard deviation units standardized at the district level.

* Difference is statistically significant at the 0.05 level using a two-sided test.

[22] We translated the student-level standard deviations into percentiles assuming that student test scores are normally distributed.

Table III.2 describes the students of the highest-performing teachers compared to those of other eligible teachers in the four districts for which we estimated value-added scores.[23] The data describe students during the 2005–2006 through 2008–2009 school years, the period to which the value-added analysis pertains. Highest-performing teachers had, on average, significantly higher proportions of white students and significantly lower proportions of African American and Hispanic students compared to other eligible teachers. Also, they had a significantly lower proportion of economically disadvantaged students measured by free and reduced-price lunch (FRL) status, and significantly lower proportions of students who had special education or limited English proficiency (LEP) status.

Table III.2. Student Characteristics: Highest-Performing vs. Other Eligible Teachers (percentages)

Student Characteristic	All	Highest-Performing Teachers (Top 20%)	Other Teachers (Bottom 80%)	Difference
Demographic				
Male	52.8	51.7	53.0	-1.3*
White	37.4	45.0	35.4	9.6*
African American	34.0	29.0	35.3	-6.3*
Hispanic	22.4	19.5	23.1	-3.6*
Economic				
Free/Reduced-Price Lunch Status	60.3	53.2	62.1	-8.9*
Academic				
Math pre-test score†	-0.06	0.16	-0.12	0.28*
English Language Arts pre-test score†	-0.07	0.12	-0.12	0.24*
Special Education Status	23.1	19.4	24.0	-4.6*
Limited English Proficiency Status	9.2	7.6	9.6	-2.0*
Sample Size (teachers)	**3,751**	**772**	**2,979**	

Source: Administrative data.

Notes: Data pertain to the subgroup of four districts whose value-added estimates were calculated by the study team. Sample size for free/reduced-price lunch status, pre-test scores and limited English proficiency status is lower because of missing data at the student level.

† Pre-test scores are standardized within a grade at the district level.

* Difference is statistically significant at the 0.05 level using a two-sided test.

b. Highest-Performing Teachers Already in Low-Achieving Schools

As mentioned in Chapter I, the TTI anticipated the possibility that some of the district's highest-performing teachers would already be serving in low-achieving schools and hence ineligible for the transfer incentive. These teachers were eligible instead for a retention stipend of $10,000 over the same two-year period, without having to apply, change schools, or be accepted by a new principal. Retention teachers were eligible for the incentive in any low-achieving schools, whether they were treatment, control, or not part of the study.

[23] Administrative data on student background characteristics was not available for the other three districts.

In the seven cohort 1 districts, retention teachers represented 12 percent of the highest-performing teachers identified by the program, or one retention candidate for every eight transfer candidates. The proportion of the highest-performing teachers already teaching in low-achieving schools ranged across the seven districts from 1 percent to 18 percent. This variation reflects differences in the underlying distribution of these teachers.[24] Just over half (51 percent) of all receiving schools had at least one retention teacher. The percentages within district ranged from 7 to 71 percent of potential receiving schools that had at least one retention teacher.

If we focus on just the teams of teachers in the study (treatment or control), we find that there were 26 high-performing teachers already in low-achieving schools. These 26 teachers were distributed across treatment groups as follows. Out of 241 teachers in treatment teams, 17 teachers (7.1 percent) were highest-performing and out of 209 teachers in control teams, 9 teachers (4.3 percent) were highest-performing. The difference is not statistically significant at the 0.05 level.

2. Identifying and Filling Vacancies

For the TTI study, a site manager designated by The New Teacher Project (TNTP) was the primary person responsible for all school and teacher recruitment. The site manager worked on both sides of the match process, both with the transfer candidates and the receiving-school principals, working to fill a set number of vacancies for each pool (elementary, middle school math, and middle school language arts), determined at the district level.

a. Outreach to Transfer Candidates

Experience from the pilot study suggested that simply informing highest-performing teachers about a $20,000 incentive for transferring to a low-achieving school might not be sufficient to encourage their participation. The TTI program relied on extensive outreach by the site managers, who served as a single point of contact for teacher candidates in each district and conducted three main recruitment activities: sending invitation letters, organizing a reception that also served as an information session, and maintaining frequent communication with teacher candidates to solicit their participation and invite them to apply and interview for specific openings.

Site managers contacted TTI candidates by email or phone at each step in the recruitment process: after sending the initial invitation letter, following the information session, as vacancies became available, and after teachers attended interviews. Where necessary (in one of the seven districts), a second information session was offered to increase the number of applicants. Conference calls were offered to answer questions posed by candidates. Teachers who did not apply to the program received ongoing communication from site managers about vacancies available through the program. Site managers typically targeted recruitment based on the grade level or subject area a teacher wanted to teach or teachers' geographic proximity to eligible vacancies. For districts that had delays in identifying eligible vacancies, teachers received update emails to inform them about the status of the process. In the pilot district, the site manager

[24] A more complete analysis of the distribution of highest-performing teachers is presented by Glazerman and Max (2011), who relied on data from many of the same school districts in the current study.

arranged for a participant who had been through the program during the previous year to serve as an ambassador to new candidates considering the program, answering questions for prospective participants.

District leaders also played an important role in recruitment. The district superintendent or human resources manager signed the initial invitation letter and attended the teacher reception to express the district's support for the program. At the reception, district staff typically discussed why the program was important and explained how it fit with existing district initiatives. Where necessary, the human resources staff sent additional emails urging unresponsive teacher candidates to apply to the program and informing them of the status of transfer opportunities at various stages.

The recruitment materials and communication from site managers relied on a variety of messages about the program. According to site managers, no one type of message was clearly more productive than another. Recruitment messages included the following:

- **Educational equity.** Recruitment materials communicated the importance of ensuring that all students in the district had access to strong teachers. The initial invitation letter noted that the program encouraged highest-performing teachers to transfer to the schools where they were most needed and could have the most profound impact. The program tag line of "Change Schools, Transform Lives" also promoted the notion that participating in the program could make a difference in the lives of students in low-achieving schools.
- **Their value as a highest-performing teacher.** Site managers communicated to teachers that the program needed teachers who had achieved past success in making learning gains with students. The invitation letter and reception recognized and congratulated teachers on their past performance. The invitation letter stated, "You have been invited to apply to the program because of your extraordinary achievements with students, your demonstrated teaching expertise, and your commitment to the children of [district name]."
- **Professional growth opportunity.** Site managers presented the program as an opportunity for professional growth and a new teaching experience. TTI was described as an opportunity to transfer their success to a new setting. While the TTI candidates had achieved success in their current schools, transferring into targeted low-achieving schools offered a new professional challenge. TTI also provided an opportunity for a teaching experience with students, teacher colleagues, and neighborhoods that potentially differed from their current setting.
- **Monetary incentive.** The recruitment materials clearly communicated that teachers would receive $20,000 over two years for teaching in the low-achieving school. This bonus was a supplement to the regular salary, did not count toward calculation of pension benefits, and was independent of their performance at the new school.

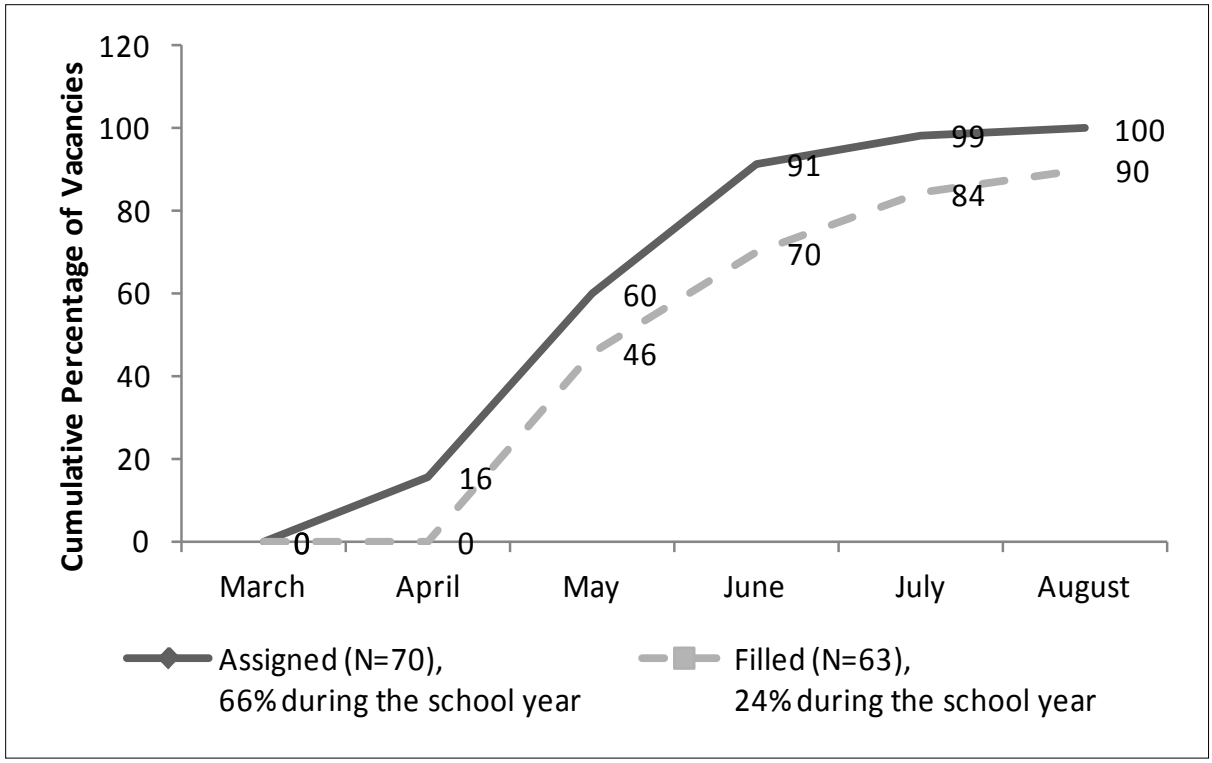
b. Outreach to Receiving-School Principals

Site managers identified vacancies in receiving schools by contacting principals on a weekly basis to learn about official and unofficial vacancies and by communicating with district human resources staff to find open positions through the formal system. Unofficial vacancies include

oral notification by a teacher of the intent to resign, retire, or transfer, even if the transition is not completed.

The site managers identified 66 pairs or groups of vacancies in total that the research team randomly assigned.[25] Because we paired vacancies to be assigned, vacancies could be assigned in blocks as they became available and as soon as a matched pair could be formed, between April and August. The bulk of vacancies were assigned and filled in May and June. Figure III.1 shows that 75 percent of vacancies were assigned and 70 percent filled in these two months. The time between randomly assigning and filling a vacancy was generally short, with site managers reporting that vacancies were filled within as few as two days of being assigned. This pattern of assignment and filling of vacancies was not different when we looked at elementary and middle school vacancies separately (see Figures C.1 and C.2 in Appendix C). Among the 70 vacancies assigned to treatment, 63 (90 percent) were filled with a TTI candidate by the end of recruitment season.

Figure III.1. Percentage of TTI Vacancies Assigned and Filled, by Month



Source: TTI program records.

The particular month of vacancy assignment and recruitment may be less important than whether assignment and recruitment took place before or after the end of the school year. Two-thirds of vacancies were assigned during the 2008–2009 school year, before the placement for fall 2009, and nearly one-quarter were filled before the end of the school year. One district had

[25] Each of the pairs or groups of vacancies was considered as a block. Random assignment was carried out by blocks, as discussed in chapter I, resulting in a total of 70 vacancies on teaching teams assigned to treatment.

15 vacancies randomly assigned a week before school ended and filled all of these positions the week after school ended. In addition, many of the other vacancies were assigned with less than two weeks remaining in the school year. There were exceptions, however. Two districts assigned and filled their vacancies after the school year ended, one due to delays in a school reorganization that prevented the program from moving forward and the other as a result of difficulty identifying placements for surplus teachers.

c. Implications for Program Costs

The steps just described have important implications for forecasting the cost of future programs like TTI. The identification and recruitment of transfer candidates, all of which happens in the six months before the transfer teachers even begin working in their new settings, constitutes the most labor-intensive part of the transfer incentive intervention and hence has major implications for cost. The other cost drivers are of course the payments to teachers, both the transfer teachers and the retention teachers (high-performing teachers already in low-achieving schools). We discuss each of these cost implications next.

The amount of effort required to identify candidates (to conduct value-added analysis and verify teacher eligibility) depends on the quality and availability of the data, and varied considerably by district. In some cases it took almost three months of intensive effort: cleaning data, merging dozens of files, reconciling anomalous data, matching ID codes, accounting for multiple courses and multiple tests, performing checks, and repeating the process for updated data when errors or omissions were found. The final step of verifying teacher eligibility took one to five days. The effort associated with recruiting teachers (transfer candidates) and receiving-school principals, from the initial information sessions to the final matchmaking and signing of transfer agreements, also varied by district and local circumstances, but typically took about five months of one-third to half-time effort on the part of a site manager. (Each site manager was responsible for up to three districts at a time and worked full time on the project).

The other major cost component is the teacher payments themselves. The cost of transfer incentives is straightforward to calculate. It is $20,000 per transfer, minus the value of pro-rated payments forfeited by teachers who leave before fulfilling their two-year commitment.[26] Policymakers can set a cap on the number of slots to fill with transfer incentives.

In addition to transfer incentives, however, future implementers must consider the cost of retention stipends for eligible highest-performing teachers who are already in low-achieving schools. Unlike transfer teachers, these teachers do not have to apply and interview to be eligible for payments, so their costs cannot be capped ahead of time. The number of such teachers is only known when the value-added analysis is completed and the teachers' current (or expected) school assignments are known for the coming year.

[26] The percentage of teachers who leave before fulfilling their commitment will be tabulated and discussed as part of a full retention analysis that will be included in the final report for this study.

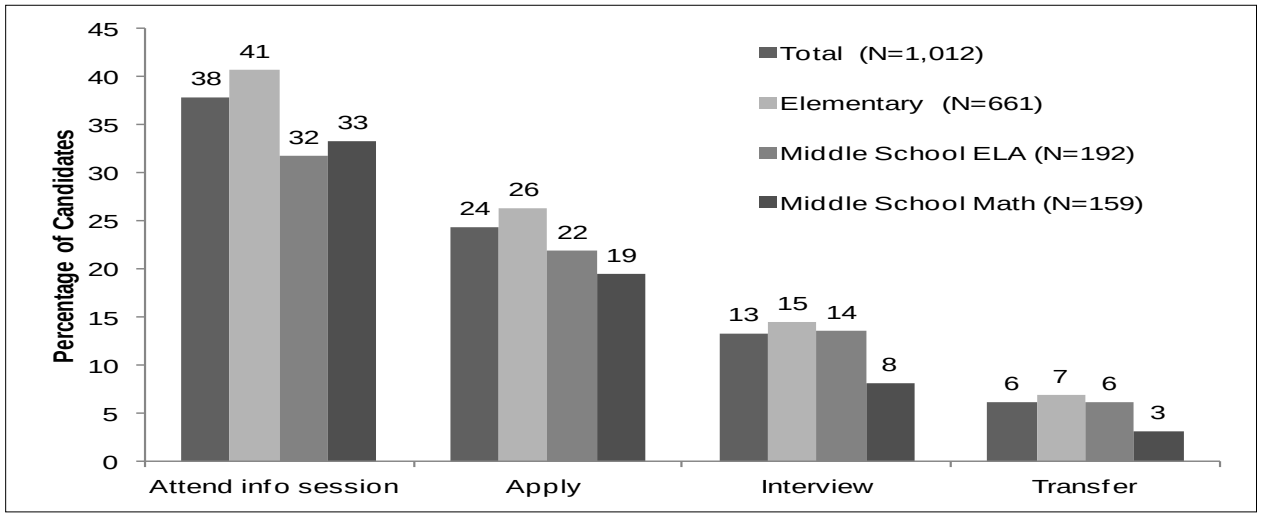
In the seven districts participating in TTI, we found that there were 133 teachers eligible for retention stipends, compared to 70 positions assigned to the TTI.[27] One reason for such a high ratio (almost two to one) is that study required identification of approximately double the number of teacher teams to form a control group and there were schools that were not participating in the study at all, but whose teachers were eligible for retention stipends because the schools were low achieving. In addition to the 70 vacancies assigned to the TTI we assigned 67 vacancies to the control group, so if we assume that 90 percent of those would be filled, the ratio for the purpose of planning future programs should be more like 1.1. That means that if a policymaker sets aside a budget of $1,000,000 to pay for transfer teacher stipends of 50 teachers, then another $550,000 should be set aside for the 55 teachers we expect to be eligible for retention stipends. None of these estimates account for teacher attrition. Teachers who leave before the two years under either type of payment will not receive the full payment amount. Data on attrition will be presented in a future report.

B. How Did Teachers and Principals React to the Transfer Incentive?

1. Take-up Rates

To gauge the response of the candidates in the Talent Transfer Initiative, we examined the rates at which candidates took part in various phases of the process ("take-up rates"), from attending information sessions to completing an application, interviewing, and ultimately transferring. Figure III.2 provides a breakdown of the take-up rates by grade span and subject using TTI program records.

Figure III.2. Take-Up Rates Among Transfer Candidates by Level in the Seven Cohort 1 Districts



Source: TTI program records.

Notes: Transfer candidates are the highest-performing teachers, ranked in the top 20 percent of value-added ranking in their pool within their district. We considered 3 pools: Elementary, middle school English /language arts and middle school math.

[27] The teaching teams assigned to treatment had 17 teachers eligible for retention stipends. The other 116 teachers eligible for retention stipends were in control teams, non-study teams in study schools, or non-study schools.

Most teachers who were offered the transfer incentive did not apply for it or even attend an information session. Forty-one percent of the elementary candidates attended the information session/reception. For middle school the attendance rates were 32 percent of ELA teachers and 33 percent of math teachers. More than one-third of the candidates who attended the information session did not complete an application for any TTI position—application rates (as a percentage of all applicants including those who did not attend the information session) were 26 percent for elementary school 22 percent for middle school ELA and 19 percent for middle school math. In the Candidate Survey, we asked the candidates if they used the information session as one of the sources to obtain information about the TTI, and if so, whether they found it useful. Forty-five percent of those who responded used the information session as a source, and most of them (98 percent) found it useful irrespective of their application status.[28]

Of those who expressed initial interest, a majority followed through to the interview stage. Fifty-four percent of applicants (13 percent of all candidates) interviewed for at least one vacancy. The other 46 percent of those who applied for a TTI position either did not follow through or were not given a chance to interview. Figure III.2 shows these numbers separately by pool (grade span and subject). Candidate Survey data suggest that of those who did interview, 85 teachers (63 percent) interviewed at one school, 27 (20 percent) interviewed at two schools, and the remaining 24 (17 percent) interviewed at three or more schools.

Data on interviews, offers, and acceptances that we obtained from the site managers provide some insight into the selectivity of the hiring process. Of the 135 candidates (13 percent of 1,012) who interviewed for at least one of the 70 available TTI vacancies, there were a total of 226 interviews attended by the candidates, because some candidates interviewed for more than one position. Thus, the average number of interviews per TTI vacancy was about 3.2. Offers were made to 74 candidates, of whom five received two offers each and one received three offers.[29] In total, 81 offers were made by TTI schools to fill the 70 TTI vacancies, or 1.2 offers per TTI vacancy and 0.4 per candidate interviewed. This means that most principals with treatment teams made offers to only one TTI candidate and to one of three they interviewed. However, principals in TTI schools could have made offers to candidates that were not TTI candidates.

More details on the hiring process from the candidates' perspective are presented in tables C.1 and C.2 in Appendix C. Candidates found most of the interviews to be informative, and said they provided opportunities to communicate their strengths, answered their questions, were conducted with genuine interest by the principal, conveyed that the interviewer was someone they could work with, and increased their desire to teach at the school. Most of the interviewees (61 percent) met one-on-one with the principal or assistant principal, and an equal percentage reported that they interviewed with other school staff. (The two types of interviews are not mutually exclusive). It was less common for candidates to give a teaching demonstration (10 percent), receive a school tour (38 percent), or meet students at the school (9 percent).

[28] The other sources used most by the candidates were internet, printed materials, and phone/email contacts with TTI program staff, 53, 56, and 77 percent, respectively. More than 90 percent of those who used these sources found them useful irrespective of their application status.

[29] The number of offers received is self-reported by candidates in the Candidate Survey.

2. Hiring Process from the Principal Perspective

Using data from the Principal Survey, we examined how the hiring rates in the treatment teams (that is, grades in receiving schools with vacancies eligible for a TTI transfer stipend), compared to those in the control teams. Hiring in control teams represents the normal hiring practice of the low-achieving schools that would be observed in the absence of the TTI. Principals were asked to report the number of applicants they considered, interviewed, and offered a TTI position to, and how many acceptances for these offers they received for each grade with a vacancy.

The average number of applicants considered per vacancy was 4.4 in treatment teams and 4.3 in the control teams (Table III.3).[30] However, these numbers are lower if schools that did not consider any applicants (or schools for which principals did not report considering any applicants) are included in the calculation. The majority of the applicants considered, 79 percent or 3.4 per vacancy in the treatment teams and 81 percent or 3.5 per vacancy in the control teams were interviewed. Offers were made to 40 percent of the candidates interviewed in the treatment teams, of which 93 percent were accepted. The corresponding numbers for the control teams were 32 percent and 96 percent. None of the differences between the treatment and the control teams was statistically significant. One mechanism by which a transfer incentive intervention is hypothesized to be effective is by making a hard-to-staff school more attractive to candidates and by providing their principals with a larger pool of candidates. These principal survey findings do not provide support for this hypothesis.

Table III.3. Hiring Rates in the Treatment and Control Teacher Teams

	Treatment	Control	Difference	P-value
Applicants Considered per Vacancy	4.35	4.25	0.10	0.92
Applicants Considered per Vacancy (including teams where no applicants were considered)	2.84	2.87	-0.03	0.98
Applicants Interviewed per Vacancy	3.44	3.46	-0.02	0.97
Applicants Interviewed per Vacancy (including teams where no applicants were interviewed)	2.25	2.33	-0.08	0.86
Offers Made per Applicant Interviewed	0.40	0.32	0.08	0.20
Offers Accepted per Offer Made	0.93	0.96	-0.03	0.46

Source: 2009–2010 Mathematica Principal Survey.

Notes: Analysis conducted at the teacher team level. Sample sizes are 34 treatment teams and 26 control teams.

[30] A recent study of all Chicago Public Schools for the 2006–07 school year, found an average of 12.7 applicants per vacancy (Engel and Jacob 2011). They reported 54.5 applicants per school and 4.3 vacancies per school, from which we calculated the applicants per vacancy.

In the Principal Survey, we also asked principals in schools with treatment or control teams about the teacher recruitment and hiring processes at the particular grade-subject teams in their schools.[31] We summarize their responses in Table III.4. Thirty-six percent of principals listed "superior teaching skills" as their most important reason for hiring teachers, while another 29 percent of principals listed this as their second or third most important reason. "Content knowledge" was deemed the most important reason for hiring by 26 percent of principals. On the other hand, no principal listed "willingness to do something extra beyond classroom teaching" and "ability and willingness to teach different grade levels" as the most important reason for hiring.

Table III.4. Teacher Characteristics That Principals with Treatment and Control Vacancies Look for When Hiring (percentage)

Teacher Characteristic	In Top Three Reasons	Most Important Reason
Superior Teaching Skills	64.8	36.4
Content Knowledge	65.9	26.1
A Passion for Teaching	27.3	11.4
Cares About Children	42.0	8.0
Willingness to Work as Part of a Team	30.7	6.8
Classroom Management Skills	45.5	6.8
Willingness to Do Something Extra Beyond Classroom Teaching	12.5	0.0
Other	8.0	4.5
Sample Size (principals)	**88**	**88**

Source: 2009–2010 Mathematica Principal Survey. The survey item was worded as follows: "What characteristics do you look for when hiring a classroom teacher?"

Notes: Means shown are for principals with both treatment and control vacancies. The differences in treatment-control means are not statistically significant at the 0.05 level. Responses are from both elementary and middle school principals.

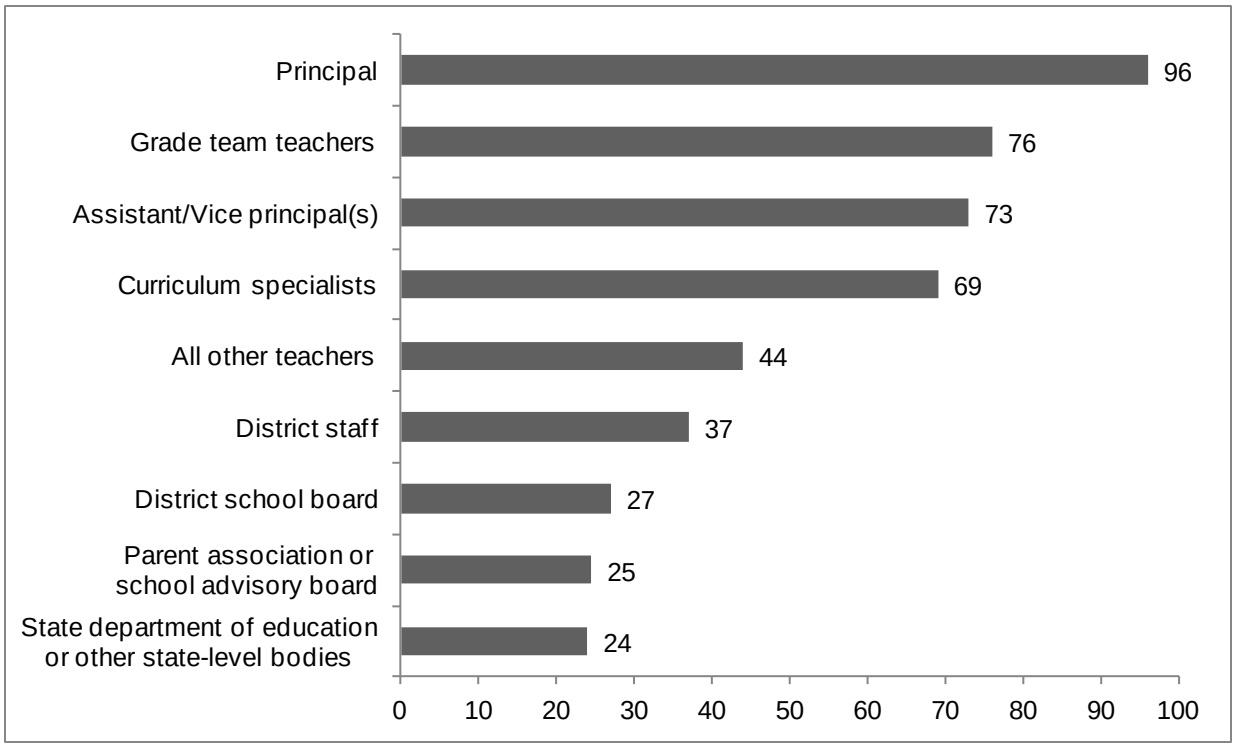
Another component of the hiring process was the influence of other groups or individuals. As shown in Figure III.3, we observed that school personnel had the most influence on hiring decisions: principals had the greatest influence (96 percent), followed by teachers teaching in the same grade (76 percent), and assistant principals (73 percent), and curriculum specialists (69 percent). Only 24 to 37 percent of principals reported groups or individuals outside of the school, including the district central office, as having influence in the hiring process.

[31] We refer to Principal Survey respondents as principals even though some principals designated an assistant or vice principal to complete the questionnaire. While principals had the opportunity to fill vacancies in treatment teams with a TTI candidate, they filled their vacancies in control teams as they normally would.

C. Which Teachers Applied to Transfer and Successfully Transferred?

Information from the Candidates Survey combined with data on whether candidates applied or transferred tells us about the transfer behavior of the candidates. Our goal is to examine why some candidates did not take the first step by applying for the transfer incentive of $20,000 and why candidates did not ultimately transfer.

Figure III.3. Influence of Different Groups on Hiring Process in Schools with Treatment and Control Vacancies (percentage)



Source: 2009–2010 Mathematica Principal Survey. The survey item was worded as follows: "How much influence does each of the groups or individuals listed below have on the process of selecting and hiring new full-time classroom teachers at your school?" For each group or individual, principals could choose "Major Influence," "Moderate Influence," "Minor Influence," or "No Influence/NA." Groups or individuals were considered to influence the hiring process if the principal indicated that they had moderate or major influence.

Notes: Means shown are for principals with treatment or control vacancies (N=90). The differences in treatment-control means are not statistically significant at the 0.05 level. Responses are from both elementary and middle school principals.

For candidates who did not apply to TTI, we explored the reasons for not applying given by these candidates themselves in the survey (Table III.5). When candidates were asked to list all of the reasons why they did not apply, being happy at their current school was most frequently listed, by 91 percent of the candidates. Commuting issues and concerns regarding not being able to return to their current school were the next two most frequently listed reasons (53 percent of the candidates), followed by concerns about not being welcome at the TTI school (43 percent), concerns about TTI school and/or neighborhood safety (36 percent), and child care and/or family-related issues (33 percent). Interestingly, 29 percent of the candidates reported that one of the reasons for not applying was their lack of confidence about being effective as a TTI teacher. Twenty-five percent reported that the stipend of $20,000 was not large enough for them to consider applying.

Table III.5. Top Self-Reported Reasons for Not Applying to TTI

	All Reasons (marked all that applied)	Most Important Reason
Factor	Percentage	Percentage
Happy at Old School	91.0	32.4
Child Care or Family-related Issues	32.5	15.2
Concern About Not Being Able to Return to Current School	52.5	6.5
Commuting Issues	52.5	6.3
Students in Receiving Schools Too Challenging	25.0	5.6
Concerns About Being Unwelcome and Not Receiving Enough Support at the TTI School	43.0	4.1
Stipend Not Big Enough	24.7	3.4
Not Confident About Self-effectiveness in a TTI School	29.2	3.4
Committed to Another School and Did Not Want to Go Back on Word	26.1	2.9
Do Not Support the Philosophy of the Program	11.6	2.9
Timing of Application Did Not Work Out	20.5	2.8
Grade Level/Subject Area of Vacancies Not Ideal	16.3	2.3
Concerns About TTI School/Neighborhood Safety	36.4	2.3
Did Not Like the Principals at the TTI Schools	5.5	1.6
Application Process Too Difficult or Too Time Consuming	10.0	1.4
Other	0.7	0.5
Sample Size	**680**	**680**

Source: 2009–2010 Mathematica Candidate Survey.

When asked about which of these reasons was the most important in deciding not to apply, being happy at their current school was cited by the highest number of candidates, 32 percent.

Child care and/or family-related issues were cited by 15 percent as the most important reason for not applying, the second-highest category. None of the other factors was mentioned as the most important reason by more than 7 percent of respondents.

We also examined the demographic, residential, and professional characteristics of candidates by their application status to better understand the influence of their observable background characteristics in their career choices. Table III.6 presents a comparison of characteristics between those who did not apply, those who applied but did not transfer, and those who transferred.

Table III.6. Characteristics of Candidates by Application Status (percentages unless otherwise noted)

Characteristic	Did Not Apply	Applied but Did Not Transfer	Transferred	All Candidates
Demographic				
Age				
25-35	20.1	32.1	24.1	22.1
36-45	26.1	24.1	40.2	26.1
46-55	30.1	26.1	26.1	28.1
55+	24.1	18.0	8.0	22.1
Male	14.0	16.0	12.0	14.0
White	68.3	58.2	44.2	64.3
Married	72.4	60.3	58.3	70.4
Have co-residing children	42.2	44.2	54.2	42.2
Residential				
Own Home	90.4	84.4	84.4	88.4
Average Commute Time				
Under 10 minutes	18.0	18.0	6.0	16.0
10-25 minutes	54.2	48.2	48.2	52.2
25+ minutes	28.1	34.1	46.2	30.1
Professional				
Base Salary (dollars)	49,621	44,270	43,726	48,065
Other Compensation (dollars)	3,571	3,193	3,380	3,479
Years of Experience in Teaching				
0 (first year teaching)	0.0	0.0	0.0	0.0
2-5 years	6.0	10.0	8.0	8.0
6-10 years	24.1	38.1	36.1	28.1
11+ years	70.3	52.2	56.2	64.3
Has a master's or doctorate degree	44.2	48.2	48.2	46.2
Has National Board Certification	24.1	24.1	22.1	24.1
Sample Size	**602**	**170**	**63**	**835**

Source: 2009–2010 Mathematica Candidate Survey.

Because many of the candidate characteristics listed above are related to one another, it may be difficult to understand which factors are still related to the probability of a candidate applying and/or transferring if we hold the other factors constant. We performed multivariate analyses to understand which factors correlate with candidates' decision to apply and transfer. In particular, we used a logistic regression with application status (whether a candidate applied or not) as the outcome and then repeated the analysis using transfer status (whether a candidate transferred or not) as the outcome.

The explanatory variables included in the regressions are the following: (1) a measure of income of the candidate, which is the base salary plus any compensation; (2) a set of personal characteristics of the candidate including gender, race, marital status, and an indicator for whether the candidate has co-residing children under age 5; (3) a set of residential characteristics, including whether the candidate owns a home and average commute time from home to current school; (4) a set of professional characteristics, including candidate's degree and National Board Certification status; and (5) a set of indicator variables summarizing the candidate's satisfaction with different aspects of his or her current school, including school leadership/policy, payments and benefits, professional environment, school environment and facility, and students.[32] We also accounted for any unobserved (by the researcher) effects at the district level influencing decisions of all candidates within districts similarly, such as district union policies or district labor market conditions. Standard errors of estimated logit coefficients account for clustering at the school level, to account for the possibility of unmeasured factors common to the same potential sending school at which multiple candidates may teach.

Relative to the teachers who did not apply, teachers who applied to TTI were different in some consistent ways. They were more likely to be African American, be unmarried, have lower income, be satisfied with their current pay and benefits, or be less satisfied with their current school policy.[33] To illustrate the magnitude, at the average level of income of $51,000, African American teachers were 20 percentage points more likely than white teachers to apply for a TTI position, all other things being equal. Also, at the average level of income of $51,000, unmarried candidates were 14 percentage points more likely to apply for a TTI position. None of the other personal, professional, or residential characteristics were significant predictors of application.

[32] The indicator variables for satisfaction were constructed from a series of aspects of a candidate's current school for which the candidate chose his or her satisfaction level on a four-point likert-type scale—very dissatisfied, somewhat dissatisfied, somewhat satisfied, and very satisfied. A candidate was assumed satisfied for an aspect if he or she were somewhat or very satisfied. Aspects were pre-grouped in the survey questionnaire to reflect satisfaction with school leadership/policies, compensation, professional environment, school environment and facility, and students and their families. We also conducted factor analysis to confirm that the items (aspects) loaded into the five pre-defined categories. The dummy variables summarizing satisfaction with the five categories were constructed as follows: a candidate was defined as satisfied for a category and was given a value of 1 if he or she were satisfied with more than 50 percent of the aspects within that category, or given a zero otherwise. We also constructed an alternative set of dummy variables using a more restrictive definition, where a candidate was defined as satisfied for a category if he or she were satisfied with all the aspects within a category. However, using this alternative set of dummy variables did not change the regression results.

[33] For all of these variables and any other variables that are reported to be significant in this section, the p-value in the regression was less than 0.05. Complete regression results are presented in Table C.3 in Appendix C. In a different specification we used base salary instead of income and the results were the same. The correlation between base salary and income is 0.91.

Next, we examined the probability of transferring for candidates who applied. Among those teachers who applied, teachers who went through the entire process and transferred to a low-achieving school were not significantly different in any of their personal background characteristics than those who did not transfer.[34] That is, once the decision to apply to the incentive program was made, none of the teachers' personal background characteristics was a statistically significant predictor of transfers.[35] However, other factors did help explain transfer decisions. Transfer candidates who were satisfied with their students during the application period, contrary to the hypothesized direction of the effect, were three times *more* likely to transfer than those who were not satisfied with their students, a statistically significant relationship.

In addition to the information on candidates' personal, professional and residential characteristics from the candidate survey, we also examined if pre-transfer student characteristics and value-added scores of candidates were related to their decisions to apply and/or transfer. Because students in the potential receiving schools are perceived as more disadvantaged, we hypothesized that candidates who have a higher percentage of disadvantaged pre-transfer students might be more willing to apply for and/or transfer to TTI positions.

However, data on student characteristics and value-added scores measured prior to the candidate transfer were available only for the four districts in which we estimated value-added scores and had student-level data. Table III.7 presents the value-added scores and student characteristics by candidates' application status for these four districts.

Table III.7. Value-Added Scores and Student Characteristics of Candidates by Application Status (percentages except for value-added scores)

Characteristic	Did Not Apply	Applied but Did Not Transfer	Transferred	All Candidates
Value-Added				
Math (score)[a]	0.23	0.23	0.20	0.23
Reading (score)[a]	0.13	0.14	0.13	0.13
Percentage in Top 10	51.3	52.6	48.6	51.3
Demographic				
Male	50.7	52.3	50.1	51.0
White	57.7	33.6	31.2	51.5
African American	20.9	31.0	30.9	23.5
Hispanic	14.6	28.5	30.9	18.2
Economic				
Free/reduced-price lunch	42.0	63.4	60.9	47.1
Academic				
Special education status	20.2	16.4	18.6	19.4
Limited English proficiency status	5.9	10.4	12.2	7.1
Sample Size	**392**	**97**	**35**	**524**

Source: Administrative data and 2009–2010 Mathematica Candidate Survey. Data pertain to a subgroup consisting of four districts that provided student-level data.

[a] Value-added scores are in student-level standard deviation units standardized at the district level.

[34] Please refer to Table III.6 for descriptive statistics on candidates' background characteristics.

[35] Complete regression results are presented in Table C.4 in Appendix C.

We examined the probability of candidates applying for the four districts where we estimated value-added score using the same multivariable approach discussed above. In addition to the explanatory variables already included, we added an indicator variable indicating whether a candidate was in the top 10 percent of the value-added ranking. We also included the percent of current students who were FRL eligible.[36] As before, unmarried candidates were more likely to apply to a TTI position. In addition, candidates with a higher percentage of disadvantaged current students and candidates who were in the top 10 percent of value-added ranking—the better among the highest-performing teachers—were more likely to apply.[37]

When examining transfer behavior focusing on these four districts, none of the background characteristics of the candidates was a statistically significant predictor of transfers except for being Hispanic, which was significantly and positively associated with transfers. Neither being in the top 10 percent of the teachers rated for value added nor having higher proportions of disadvantaged students measured by FRL status made candidates more likely to transfer.

D. Where Did Transfers Come From?

The goal of a program like TTI is to help low-achieving schools by recruiting strong teachers from schools that are not low achieving. Some types of transfers may serve that goal better than others. In designing the program, it was necessary to designate all schools in the district as being in one of two groups: low achieving (potential receiving schools) or non-low achieving (potential sending schools). As described in Chapter II, school percentile ranks were calculated for elementary and middle schools separately within each district, from school composite scores in reading and math in the prior year or years to the school selection. For the analysis below, we used ranking in school average scores in 2008–09. Establishing a discrete and somewhat arbitrary dividing line in the distribution between the groups means that it is possible for a teacher to transfer from a school just above the threshold to one just below it. We call these moves lower-contrast transfers because the difference in achievement ranking between the sending and receiving schools is small. The transfer incentive might be counterproductive in the case of lower-contrast transfers if the sending school is itself in need of strong teachers and has difficulty filling the vacancy created by the transfer.

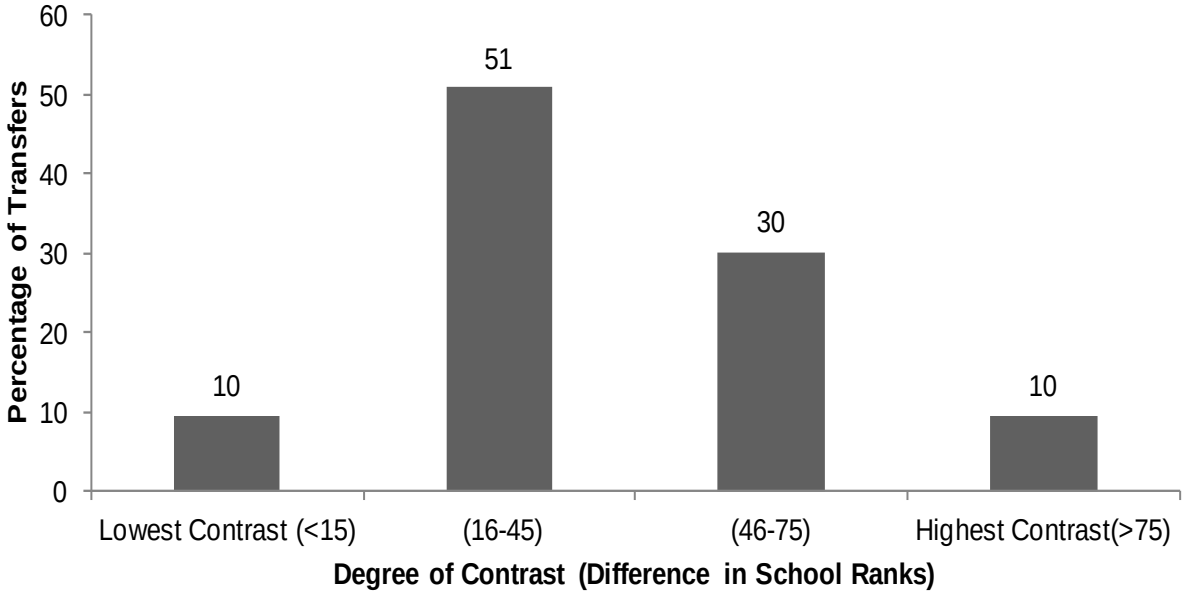
We grouped the transfers by the degree of contrast, measured as the difference in the rank between the sending school and the receiving school for a given transfer. The maximum degree of contrast would be a transfer from the highest-achieving school in the district to the lowest-achieving school, a difference of 100 percentile points.

[36] Percent African American and percent with limited English proficiency are two other measures of student disadvantage, which we did not include in the regression because of their high correlation with percent FRL status, 0.56 and 0.45 respectively.

[37] Complete regression results pertaining to the subgroup consisting of the four districts that provided student-level data are presented in the last two columns of Tables C.3 and C.4 in Appendix C.

The average contrast in school achievement rank is about a 42 percentile points, with a median contrast of 40 percentile points. Specifically, the average sending school was ranked in the 60th percentile, where the 100th percentile is the highest achieving school in the district, and the average receiving school was ranked in the 18th percentile. This means that, on average, transfer teachers did not move from schools just above the threshold of low achieving defined for this study.[38] Sixty-one percent of the transfers involved highest-performing teachers moving between schools that were ranked within 45 percentile points of each other in the rank distribution. On the other hand, 40 percent of the transfers involved highest-performing teachers moving between schools that were more than 45 percentile points apart. Figure III.4 summarizes these contrasts for the 63 teachers who successfully transferred to low-achieving schools.

Figure III.4. Types of Transfer by Achievement Ranks



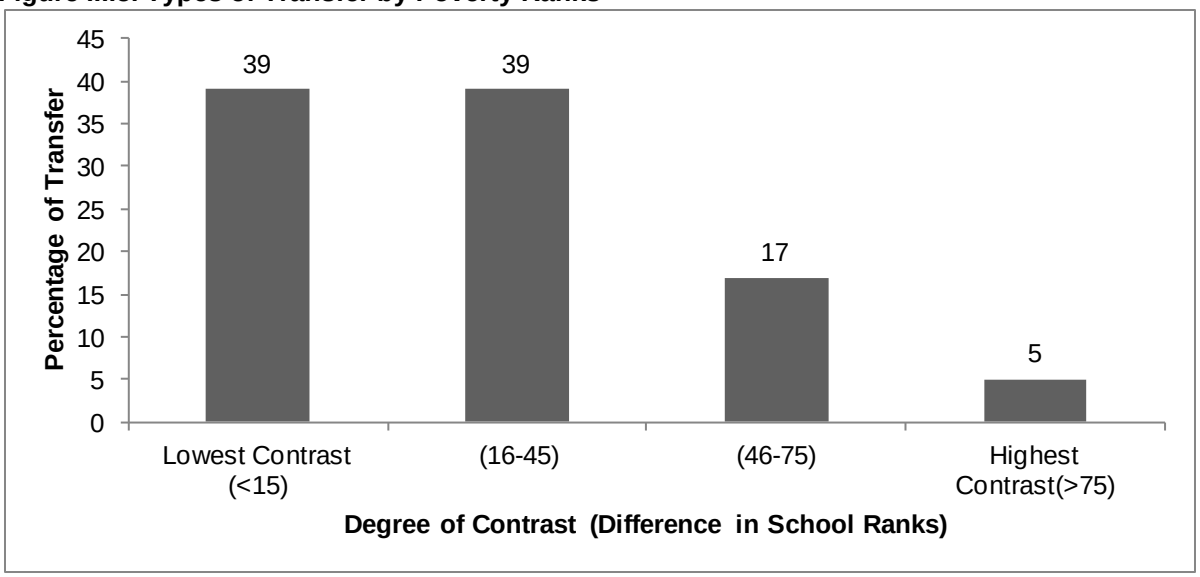
Source: Administrative data and TTI program records (N=63).

[38] Achievement rankings include exempt schools. The achievement rank of the highest-achieving potential receiving school ranged across districts and pools from 27th to 42nd.

Although the poverty status of schools was not taken into account when identifying potential receiving schools, it is still informative to examine the contrast in the percentile rank positions based on poverty status (as measured by percentage of students eligible for FRL) of schools that transfer teachers left compared to the ones to which they moved. The average contrast in poverty rank was about 25 percentile points, with an average sending school percentile rank of 55 and an average receiving school rank of 18.[39] Seventy-eight percent of transfer teachers moved between schools that were within 45 percentile points of each other based on poverty status. Figure III.5 summarizes the percentage of transfers in some of the contrast categories, based on school poverty status ranking.

As one additional way of comparing the circumstances of the transfer candidates' original schools with those to which they transferred, we compared the candidates' students before and after the transfer. This is relevant to the question of whether match effects are likely to play a role in determining the impacts of TTI.

Figure III.5. Types of Transfer by Poverty Ranks



Source: Administrative data and TTI program records (N=63).

[39] We ranked schools in descending order of poverty status as measured by the percentage of students eligible for FRL. Therefore, schools with higher poverty status or higher percentage of students eligible for FRL have a lower rank.

Because this analysis uses detailed data from before the candidates transferred, we focus on the subset of four districts that provided such data. Table III.8 shows the average student characteristics before and after transfer for the 33 transfer teachers who came from these four districts. The transfer teachers moved to locations where they would be teaching a lower percentage of white students, higher percentages of minority students, and a higher percentage of FRL-eligible students. The differences were statistically significant except for percentage African American (p-value = 0.091).[40] The transfer teachers moved to classrooms with a significantly lower percentage of special education students, a difference of 8 percentage points, although we could not distinguish from the district data how much of that difference could have been due to a reduction in the percentage of gifted and talented students, who were labeled by two districts as special education. Those districts did not provide sufficient data to allow us to exclude gifted students from the special education population.

Table III.8. Characteristics of Selected Transfer Teachers' Students Before and After Transferring

Characteristic of Average Student (percentages unless noted)	In Sending Schools 2007–2009	In Receiving Schools 2009–2010	Difference (Receiving Minus Sending)	P-value
Race/Ethnicity				
White	30.1	12.1	-18.0*	0.002
African American	32.2	40.1	7.9	0.091
Hispanic	30.8	41.8	11.0*	0.001
Low Income (percent free/reduced-price lunch)	63.6	89.3	25.7*	0.000
Special Education[a]	19.0	11.2	-7.9*	0.032
Average Reading Score[b]	-0.11	-0.39	-0.28	0.021
Average Math Score[b]	-0.06	-0.47	-0.41	0.000

Source: Administrative data.

Note: Data pertain to a subgroup consisting of four districts that provided student-level data. N = 33 teachers who transferred in the four districts and for whom detailed student data were available. Due to missing data the sample size was 26 teachers for FRL, 25 for reading scores and 23 for math scores. Not all teachers taught both math and reading.

[a] The special education category in two of the four districts includes gifted students.

[b] Average reading and math scores are given in fraction of a standard deviation computed within district and grade

* Difference is statistically significant at the 0.05 level using a two-sided test based on the teacher sample.

[40] A p-value is the probability of observing the result given that the null hypothesis is true. If the p-value is less than 5 percent then we reject the null hypothesis.

Test scores differences between transfer teachers' sending and receiving schools were also statistically significant for this group. The average student in the transfer teachers' classrooms scored 0.11 standard deviations below the district average in reading, placing them in the 46th percentile. The same teachers' students in the schools to which they transferred had scored 0.39 standard deviations below the district average, placing them in the 35th percentile. For math, the differences were -0.06 standard deviations (48th percentile) and -0.47 standard deviations (32nd percentile).

IV. AFTER THE TRANSFER PROCESS: PLACEMENT RESULTS AND INTERMEDIATE IMPACTS

This chapter focuses on the teacher teams that are affected by the transfer program, contrasting those teams having vacancies into which the highest-performing teachers had been randomly assigned to be eligible for the transfer incentive ("treatment") and those having vacancies that were randomly assigned to be filled in the usual way ("control"). We first describe the control group, to characterize the usual ways of filling vacancies and thereby understand who would have filled the vacancy had the district not implemented a transfer incentive. We then describe the treatment group and compare the two groups. The second part of the chapter compares treatment and control teachers along a number of dimensions meant to capture intermediate impacts of the program. Specifically, we hypothesized that the program may change the assignment of students to teachers, the allocation of other resources within the school, or the assignment of special responsibilities to teachers.

A. Who Filled the Vacancies?

Understanding how vacancies were filled helps us understand teacher teams, which form the basis for the current study. When we estimate the impact of TTI on student achievement and other outcomes (in a forthcoming report), our main approach will be to compare the outcomes for the teacher teams (grade-subject teams within a school) assigned to treatment to the outcomes for teams assigned to the control group. The resulting treatment-control differences are unbiased estimates of the impact of the intervention.

However, in order to interpret those impact estimates, we need to understand how the vacancies on each teacher team were filled. One might expect that the treatment group vacancies would be filled by teachers identified through the TTI process described in Chapter III, while the control group vacancies would be filled by hiring new teachers into the profession. In reality, vacancies can and were filled through a variety of means. As shown below, although most control group vacancies were filled through new hires or transfers from other schools, several were filled by moving teachers from another grade or subject within the school—and in some cases, the position was lost altogether because of declining enrollment. At the same time, while most treatment group vacancies were indeed filled through TTI, some were filled outside the program. The details are discussed next.

1. Control Group Vacancies

A transfer incentive strategy like TTI is intended to improve the quality of teachers filling new vacancies in low-achieving schools; therefore, it is important to understand how those vacancies would have been filled if there had been no intervention. Before conducting the study, we hypothesized that many such vacancies would be filled by those entering the teaching profession, but some could be filled in other ways. Teachers might transfer from other schools, be hired away from other districts, or might simply move from another grade level within the same school.

We identified 85 potential receiving schools whose principals indicated they had at least one vacancy eligible for the study, for a total of 137 vacancies. Forty-nine schools had one vacancy

but 26 schools had two, six schools had three, four schools had four or more eligible vacancies. After randomly assigning teacher teams containing those vacancies to either a treatment group that could hire through the TTI or a control group that could not, we followed the control group to learn more about the business-as-usual condition. We assigned teams with 67 vacancies to the control group and teams with 70 vacancies to the treatment group.

a. Identifying Teachers Who Filled Control Group Positions

In order to describe the teachers who filled control vacancies, whom we refer to as "control focal teachers," we first had to identify them. Although some school principals told us directly who filled the position, most did not. Our method instead relied primarily on the responses to the teacher survey. From the survey data we learned which respondents were new to the grade-subject team in 2009—but not all teachers completed the survey and in some cases there was more than one teacher who could have been the focal teacher.

The above teacher and principal reports were not always complete and did not always agree, so we reconciled the discrepancies and used the best available information, coding uncertain cases as we encountered them.

Because there were ambiguities, we used two different definitions of focal teacher: one using a selective rule and the other using an inclusive rule. The selective definition only classifies teachers that can be linked unambiguously to the study vacancy based on information provided by principals or teachers' self-reports. In some cases, we were able to identify the teacher that filled the study vacancy based on principal reports. This was the cleanest method of identifying control focal teachers and is consistent with the identification of treatment focal teachers. If more than one teacher was new to the team and the principal did not identify the teacher that filled the study vacancy, we designated all new teachers as focal teachers and assigned each a proportional weight that sums to the number of vacancies in the team. If we did not have any information from principals or teachers about who filled the vacancy on a team, that team did not have a focal teacher identified under the selective rule.[41]

For teams without focal teachers identified under the selective definition, the inclusive definition classifies at least someone in every control team, even if there was limited evidence of the person's likelihood of being the true focal teacher. In cases where we could not determine which teacher was the focal teacher, we included all possible candidates and assigned each of them a proportional weight that sums to the number of vacancies in the team. The sum of the weights represents the number of vacancies for which we had reliable data: 38 vacancies with the selective definition and 48 vacancies with the inclusive definition.

[41] Seventy-two percent of control teams had at least one focal teacher identified under the selective rule. All control teams except for one had at least one focal teacher identified under the inclusive definition. There was one control team for which the principal provided specific information about the teacher who was hired to fill the study vacancy, but that teacher did not appear in the teacher rosters collected from the school or in the background survey frame.

We focus most of our focal teacher analysis on the selective definition because it provides the most accurate description of the counterfactual, but it is important to recognize the tradeoff between these definitions. The selective rule has fewer ambiguous cases, but leaves more teams with missing data. The inclusive rule identifies at least one teacher for nearly every team, so there is less missing data, but can include teachers who were not in fact new to the team. We conducted the same analyses using the inclusive definitions and include the results in Appendix C.

b. Teachers Who Filled Control Group Positions

Table IV.1 shows that the control group vacancies were filled by a combination of new hires, transfers in, and within-school reassignments. The table also shows the number of positions that were not filled. This breakdown illustrates what would have happened to teaching vacancies in low-achieving schools in the absence of TTI. Fifteen (22 percent) of the 67 control positions were filled with new hires, with most of those (9 of the 15 hires) being new to the profession. Another 13 (19 percent) were transfers from other schools, and 18 (27 percent) moved from another position within the school. Most of these within-school moves were the result of moving the vacancy to another grade, so there were more teachers who may have been new to the school but who were not included in the study because we focused on the grades randomly assigned. Another six positions were filled by individuals whose background we could not ascertain because they did not respond to the Teacher Background Survey. An additional seven positions (10 percent) were lost because of a drop in student enrollment, an increase in class size, or because the teacher who was leaving to create the vacancy instead returned to his or her position. Finally, there were eight vacancies (12 percent) whose ultimate status could not be determined unambiguously. In these cases, there were so many teachers within a grade who did not complete a survey that we could not determine which teacher filled the vacancy and there was also insufficient information from the principal to determine the outcome of the hiring process.

Table IV.1. How Study Schools Filled Their Vacancies in the Absence of a Transfer Program (Control Group Only)

Final Status of the Vacancy	Number	Percentage
Positions Filled		
New to teaching	9	13.4
New hire (new to the district or new to teaching)	6	9.0
Transfer from other school	13	19.4
Transfer from another grade	18	26.9
Unknown origin/uncertain	6	9.0
Position Lost, Transfer Cancelled, or Layoff Rescinded[a]	7	10.4
Unknown Status[b]	8	11.9
All Vacancies	67	100.0

Source: 2009–2010 Mathematica Teacher Background Survey, 2009–2010 Mathematica Principal Survey, 2011 principal follow-up interviews.

[a] Teachers whose transfers out of the study school were cancelled or whose layoffs were rescinded were treated as the focal teacher for this study.

[b] These are teaching teams whose vacancy was not filled, or where the focal teacher was not identifiable.

The characteristics of control focal teachers are summarized in Table IV.2. While most teachers (58 percent) who filled the control group vacancies were new to the school, not all were new to the profession.[42] In fact, an equal 58 percent reported being in at least their sixth year of teaching. Twenty-one percent held a master's degree and 12 percent held National Board certification, an advanced teaching credential that requires a lengthy application process and demonstration of mastery through a portfolio and other materials. The experience profile of these teachers reflects the fact that many (see Table IV.1) were experienced teachers who simply moved from elsewhere in the school or district and were not hired out of the beginning teacher pool.

The demographic characteristics of control focal teachers are also shown in Table IV.2. Eighty-four percent of control focal teachers were female, 55 percent were white, and 62 percent were married. The average age of the control teachers was 37 years. Characteristics of this group are shown using the inclusive definition as well, for perspective, since the designation of focal status was uncertain for so many teacher teams. The inclusive definition allows us to include all teams in the control group with survey respondents, although it may not be as accurate as the selective definition. Of course, the selective definition may not be as representative as the full sample, so there is a tradeoff between the two approaches.

Table IV.2. Characteristics of Teachers Who Filled Control Vacancies (percentages)

Characteristic	Control Focal (selective definition)	Control Focal (inclusive definition)
Professional Background		
Years of Experience in Teaching (average years)	8.0	9.2
Years of Experience in Teaching (percentages by category)		
1 (first year teaching)	21.1	16.6
2-5 years	36.8	34.8
6-10 years	17.1	17.9
11+ years	25.0	30.8
First Year in the School	57.9	45.6
First Year in the Grade	62.2	51.4
Has Regular Certification for Grade/Subject Taught	92.1	91.2
Has a Master's or Doctorate Degree	21.1	26.1
Has National Board Certification	11.8	11.0
Personal Background		
Female	84.2	82.3
Race/Ethnicity		
White, non-Hispanic	55.3	50.7
African American, non-Hispanic	25.0	31.6
Hispanic or Latino	17.1	14.5
Average Age (years)	37.1	38.2
Married or Living with a Partner	61.8	62.1
Homeowner	51.3	54.3
Sum of Weights (number of vacancies)	38.0	48.2
Sample Size (number of respondents)	**41**	**67**

Source: 2009–2010 Mathematica Teacher Background Survey.
[a] Results are weighted to account for the possibility that more focal teachers were identified in a grade team than there were vacancies identified. The sum of weights equals the number of vacancies being described.

[42] These results describe the selective definition, which includes fewer "false positives" (that is, teachers who were not filling the vacant position submitted for the study).

Table IV.3. How Study Schools Filled Their Vacancies Using Transfer Program (Treatment Group only)

How Vacancy was Filled	Number of Vacancies	Percentage
Filled with TTI Candidate	63[a]	90.0
Filled Outside TTI	4	5.7
Position Lost or Transfer Cancelled	3	4.3
All Vacancies	70	100.0

Source: 2009–2010 Mathematica Teacher Background Survey, 2009–2010 Mathematica Principal Survey, 2011 principal follow-up interviews.

[a] Three TTI candidate were placed in a different grade in the school than the originally designated vacancy. One TTI candidate transferred but left shortly after the start of the year.

2. Treatment Group Vacancies

For the most part, vacancies assigned to the intervention were filled with TTI candidates. Sixty-three out of 70 positions (90 percent) had a successful transfer, although one of those 63 teachers left shortly after the start of the school year. Three others were subsequently moved to a different grade within the same school and are not included in our analysis of treatment focal teachers below. Of the seven positions assigned to the treatment group that were not filled by a TTI teacher, three positions were lost because of enrollment declines or because teachers were recalled from a layoff notice that had created the vacancy; the remaining four were simply unable or unwilling to select a match with TTI candidates and the principal instead hired outside that pool. Table IV.3 summarizes the status of the vacancies in teacher teams assigned to the treatment group, whether or not they were filled by a TTI transfer, and Table IV.4 presents the characteristics of the teachers who filled those vacancies and responded to the teacher survey. The teachers in the treatment group all had at least two years of experience and most (51 percent) had at least 11 years of experience. The difference in experience level between treatment focal and control focal teachers (almost five years) was statistically significant.[43] Other treatment-control differences are also pointed out in Table IV.4.

B. How Did Schools React to a TTI Transfer?

When policymakers introduce bonuses for some teachers and not others, there could be positive or negative impacts on school climate. Positive impacts could come from the fresh ideas and insights that a highest-performing teacher might bring. Negative impacts could come from resentment and morale problems that lead to a breakdown in trust and collaboration. Once the teaching staff was in place, the school leaders faced several decisions regarding the degree of collaboration, assignment of students to teachers, allocation of resources like mentoring, and the deployment of teachers in roles such as mentoring and other leadership positions. The TTI could have an impact on student achievement and teacher retention through any of these routes, as discussed in the logic model in Chapter I.

[43] Had the inclusive definition been used, the difference would have been closer to four years (3.7) but also statistically significant.

Table IV.4. Characteristics of Teachers Who Filled Treatment and Control Vacancies (percentages)

Characteristic	TTI Transfers	All Treatment Focal	All Control Focal [a]	Treatment-Control Difference	P-value
Professional Background					
Years of Experience in Teaching (average years)	13.5	12.9	8.0	4.9*	0.003
Years of Experience in Teaching (percentages by category)					
1 (first year teaching)	0.0	0.0	21.1	-21.1*	0.002
2-5 years	6.8	11.1	36.8	-25.7*	0.004
6-10 years	39.0	38.1	17.1	21.0*	0.017
11+ years	54.2	50.8	25.0	25.7*	0.007
Has a Master's or Doctorate Degree	49.2	47.6	21.1	26.6*	0.004
Has National Board Certification	24.6	23.0	11.8	11.1	0.141
Transferred via TTI	100.0	95.0	0.0	95.0*	0.000
Personal Background					
Female	89.5	88.5	84.2	4.3	0.553
Race/Ethnicity					
White, non-Hispanic	45.6	45.9	55.3	-9.4	0.363
African American, non-Hispanic	29.8	31.1	25.0	6.1	0.506
Hispanic or Latino	17.5	16.4	17.1	-0.7	0.925
Age (years)	42.8	42.3	37.1	5.2*	0.022
Married or Living with a Partner	61.0	60.3	61.8	-1.5	0.879
Homeowner	84.7	82.5	51.3	31.2*	0.001
Sample Size (number of teachers)	**59**	**63**	**41**		

Source: 2009–2010 Mathematica Teacher Background Survey.

[a] The control focal teachers are those identified using the "selective" definition.

* Difference between treatment focal mean and control focal mean is statistically significant at the 0.05 level using a two-sided test. Test was only conducted for the "All Treatment" group.

1. Did Transfer Incentives Affect Collaboration?

We did not find evidence of statistically significant impacts of transfer incentives on principals' opinions of the degree of collaboration, trust, and sharing of ideas within grade teams. Table IV.5 shows how principals rated the treatment and control group teacher teams on three dimensions of school climate, each measured on a five-point scale.[44] The level of collaboration was measured from 1 ("highly independent") to 5 ("highly collaborative"). Principals rated both the treatment and control teams between a 3 and 4 on average. When we asked about the degree of trust and mutual respect teachers had for one another, the average rating was 3.9 and 3.8 for the treatment and control groups, respectively, with the difference being statistically insignificant. This measure was scaled from "no extent" to "a great extent." When we asked principals if their teachers seek ideas from one another, we also found scores that averaged just below 4 (3.8 for treatment and 3.9 for control, another statistically insignificant difference).

Table IV.5. Principal Reports on Team Climate

Climate Measure (on a 5-point scale from "no extent" to "a great extent")	Treatment Mean	Control Mean	Difference	P-value
Current Levels				
Level of collaboration	3.7	3.6	0.2	0.405
Degree of trust and mutual respect	3.9	3.8	0.1	0.658
Teachers seek ideas from one another	3.8	3.9	-0.1	0.696
Change from Prior Year				
Level of collaboration	1.1	0.6	0.5	0.135
Degree of trust and mutual respect	1.1	0.6	0.4	0.104
Teachers seek ideas from one another	1.0	0.8	0.1	0.576

Source: 2009–2010 Mathematica Principal Survey.

Note: N = 55 treatment teams, 54 control teams for current levels; N = 48 treatment teams and 44 control teams for change from prior year (excludes new principals).

We asked about change over time and found that principals who had been in the same school the prior year, regardless of treatment status of the teams they were describing, reported higher levels of collaboration, trust, and idea-sharing, by about 1 point on each of the five-point scales comparing spring 2010 to spring 2009, before the implementation of TTI. We also found no significant treatment-control differences when we focused on elementary schools only or when we collapsed the five-point scale, comparing the percentages of respondents who answered either 4 or 5.

[44] The survey items are worded as follows:

"On a scale of 1 to 5 where 1 is 'Highly independent' and 5 is 'Highly collaborative' how would you rate the level of collaboration among teachers in grade X?"

"On a scale of 1 to 5 where 1 is 'Little or no extent' and 5 is 'Great extent' how would you rate the extent to which teachers in grade X trust and mutually respect one another?"

"On a scale of 1 to 5 where 1 is 'Little or no extent' and 5 is 'Great extent' how would you rate the extent to which teachers in grade X seek ideas from one another?"

Taken at face value, these findings suggest that there was no evidence of a breakdown of morale, nor was there a significant impact in the way that teachers worked together (or if these effects were present, then they canceled each other out). One should interpret the findings with caution, however, because the survey responses represent subjective opinions and possibly reflect respondents' tendency to offer socially desirable responses.[45]

2. How Were Students Assigned?

One type of resource allocation effect is the impact of a transfer incentive on the assignment of students to classrooms. Under normal circumstances, the school principal or teacher team may decide to assign students strategically such that the newly hired teacher works with the less challenging students. Alternatively, the new teacher may be assigned *more* challenging students, or students who are basically the same as those of their peer teachers. The control group experiences, particularly the differences between the focal teachers and non-focal teachers, tell us about the outcome of this assignment process in the absence of TTI. The question for this study is whether there was an impact on differences between focal and non-focal teachers along any key dimensions of student background. To answer this, we examined focal versus non-focal differences and compared those differences between treatment and control teams.

We examined student assignment differentials using three separate data sources: (1) administrative data on student characteristics; (2) teacher perceptions as measured by self-reports on the Teacher Background Survey; and (3) principal perceptions, from the Principal Survey. The administrative data provide objective information on a few easily observed traits like prior achievement and income, proxied by eligibility for free or reduced-price lunch (FRL). The teacher survey data rely on opinions, but allow us to capture differences not only in demographic characteristics but in students' behavioral challenges. The principal reports are also subjective, but they allow us to focus specifically on the assignment process and allow respondents to tell us directly how they intended to assign students.

a. Student Characteristics

We compared the prior achievement and demographic background characteristics of students assigned to focal versus non-focal teachers separately for treatment and control teams. We computed focal versus non-focal differences in average characteristics at the team level and then presented their distribution in Figure IV.1 for prior math scores, Figure IV.2 for prior reading scores, and Figure IV.3 for percent FRL. The test score differences are reported in standard deviation units, so a difference of 0.25 is one-quarter of a standard deviation. This size difference would separate, for example, a student at the 50th percentile from one at the 40th percentile (ten percentile points).

[45] Moreover, responses by principals to closed-ended questions about highly complex phenomena such as collaboration and teamwork within teaching teams can only be used to measure stark differences between treatment and control groups and are not intended to portray the range of interactions in a school from all perspectives.

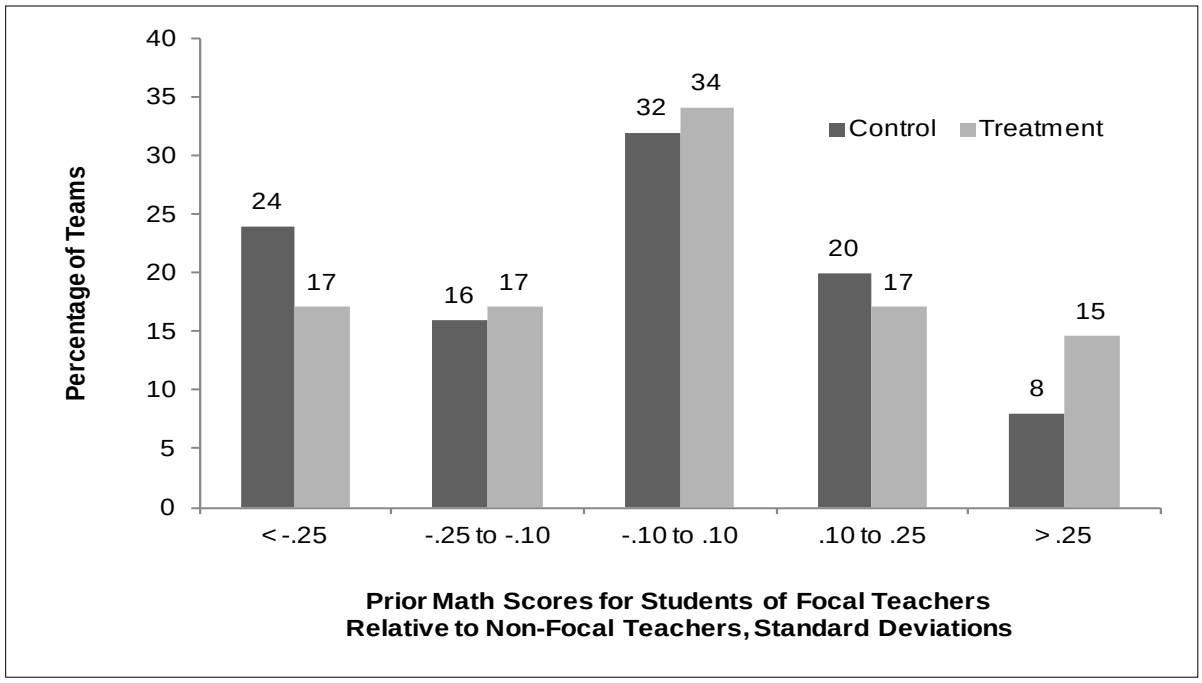
The figures illustrate two things. First, they show the degree to which "new" teachers (focal teachers, those who filled the vacancies) were assigned different students than their peers. Taller bars on the left side of the graph (less than -0.10 standard deviations or greater than 5 percent FRL) imply that they were assigned more disadvantaged students than their peer teachers. Taller bars on the right side of the graph (greater than 0.10 standard deviations or less than -5 percent FRL) imply the opposite, that focal teachers were assigned less disadvantaged students. Tall bars in the middle category (-0.10 to 0.10 standard deviations or -5 to 5 percent FRL) suggest that there was no differential assignment of students.

Second, the graphs show the extent to which this possibility of differential student assignment differed between treatment and control groups. If TTI had no impact on the way in which students are assigned to teachers, then each pair of adjacent bars in Figures IV.1 through IV.3 would be the same height. If the lighter bar (representing the control group) were lower than the darker bar on the left side of the graphs and higher than the darker bars on the right side of the graphs, then it means that transfer teachers were assigned less disadvantaged students in terms of prior achievement or percent FRL than their peers, relative to control new hires.[46]

None of the treatment-control differences were statistically significant. Figure IV.1 shows that focal teachers in 24 percent of control group teams had students with average prior math scores that were 25 percent of a standard deviation or more lower than the scores of the students assigned to non-focal teachers. In another 16 percent of control group teams the focal teachers were assigned students whose prior scores were 10 to 25 percent of a standard deviation lower than those of the students assigned to the non-focal teachers, with nearly one-third having been assigned students that were within 10 percent of a standard deviation above or below the mean for their peers' students. An additional 20 percent had been assigned higher-scoring students (10 to 25 percent of a standard deviation higher) and, finally, in 8 percent of the teams the focal teacher's students scored higher than their peers' by at least 25 percent of a standard deviation. This represents a range of outcomes, from assignment mechanisms that give focal teachers students who were lower achieving, higher achieving, and similarly situated based on prior test score results. Teacher teams in the treatment group demonstrated a similar range of outcomes, with 17 and 15 percent in the extreme categories and 34 percent in the middle category (with plus or minus 10 percent of a standard deviation differential).

[46] We use the terms "transfer teachers" and "control new hires" to describe treatment focal and control focal teachers, but it should be noted that these terms are approximate. Treatment focal teachers were not always transfers and control focal teachers were not always new hires, as shown earlier in this chapter.

Figure IV.1. Are Focal Teachers Assigned Students with Lower Math Achievement Than Are Their Peers? Results by Treatment Status

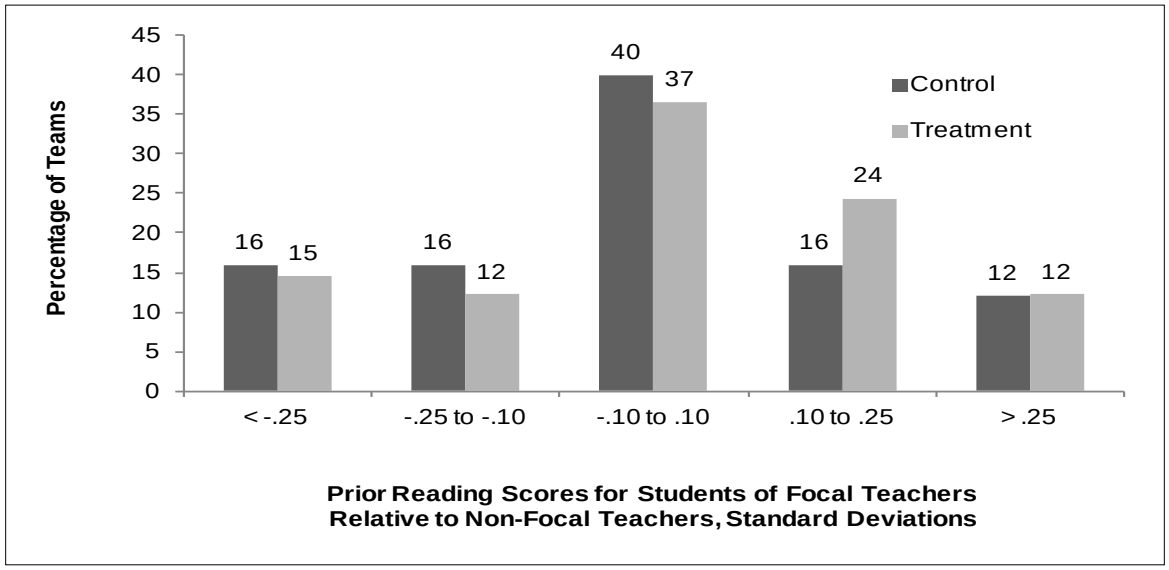


Source: Administrative data.

Note: N = 41 teams in the treatment group and 25 teams in the control group. Distributions are not significantly different based on a Pearson's chi-square test of independence.

Figure IV.2 shows the same phenomenon calculated for prior reading test scores. Here we see 40 percent of control teams and 37 percent of treatment group teams with differentials in the –0.10 to +0.10 range. The differences in distributions are not statistically significant.

Figure IV.2. Are Focal Teachers Assigned Students with Lower Reading Achievement Than Are Their Peers? Results by Treatment Status

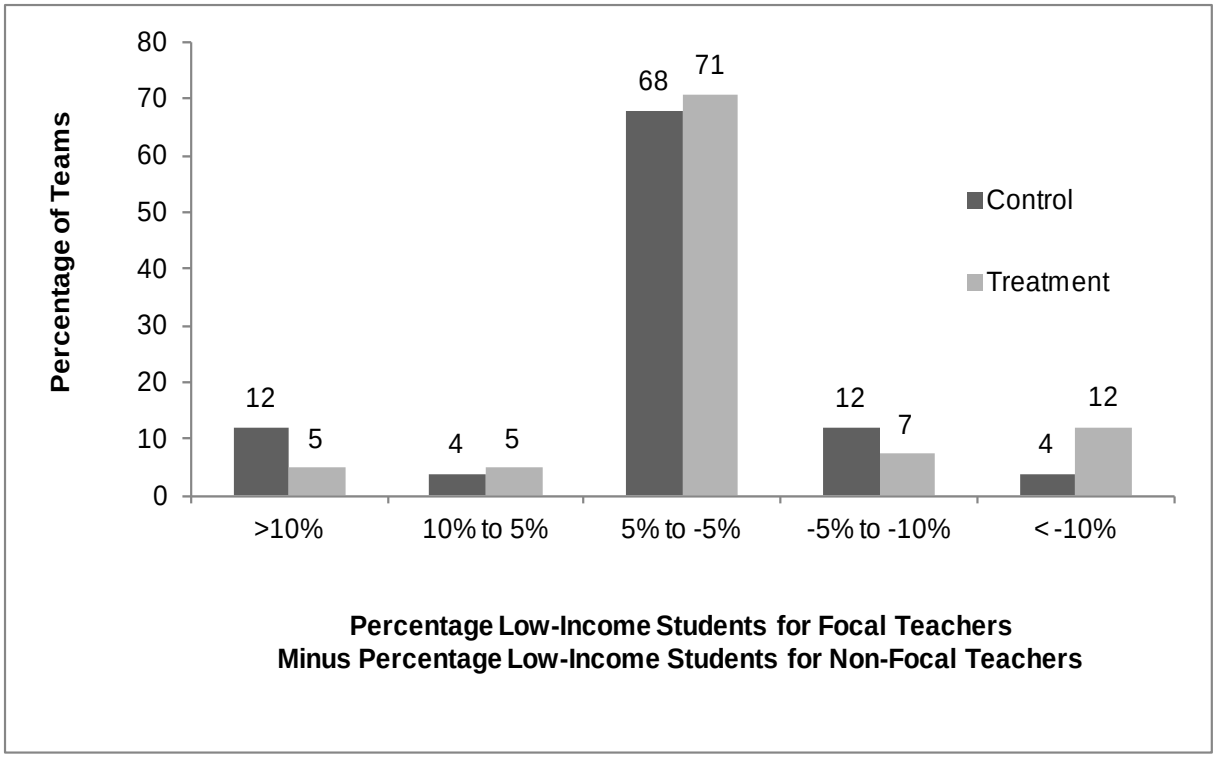


Source: Administrative data.

Note: N = 41 teams in the treatment group and 25 teams in the control group. Distributions are not significantly different based on a Pearson's chi-square test of independence.

In Figure IV.3, the distribution of focal/non-focal differences in treatment and control groups in terms of percentage of FRL students does not differ statistically between treatment and control groups. More than two-thirds of both groups had differentials that were within five percentage points (68 and 71 percent for control and treatment, respectively).

Figure IV.3. Are Focal Teachers Assigned Fewer FRL Students Than Are Their Peers? Results by Treatment Status



Source: Administrative data.

Note: N = 41 teams in the treatment group and 25 teams in the control group. Distributions are not significantly different based on a Pearson's chi-square test of independence.

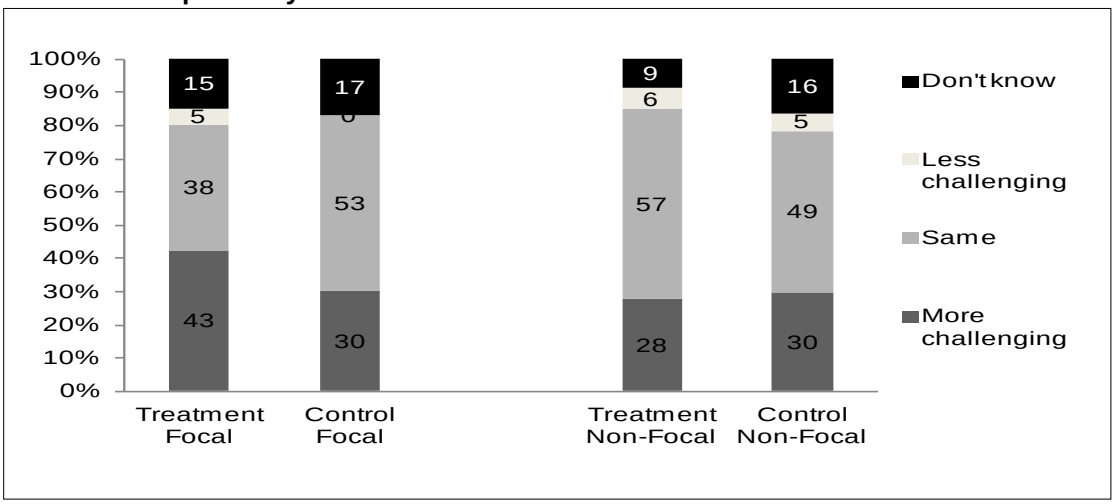
We constructed similar figures demonstrating the distributions for differentials in terms of percentages of students who were English language learners, receiving special education services, and belonging to certain race/ethnic categories (see Appendix C, Figures C.3–C.7). As with the FRL results, the differences in distributions were not statistically significant for any of these characteristics.

b. Teacher Perceptions

While prior math and reading scores provide objective information about students' academic abilities, teacher perceptions can capture more nuanced, albeit subjective, appraisals of the academic abilities of the students. Teachers were asked if they felt that their students were more challenging, less challenging, or equally challenging in terms of academic ability than the students of their peers in the same school, grade, and subject (non-focal teachers).[47] We repeated the question in terms of behavior. The results are summarized in Figures IV.4 and IV.5, which show the percentages of treatment and control teachers divided into focal and non-focal teachers.

Figure IV.4 shows the percentages of teachers in each group (treatment/control and focal/non-focal) who found their own students more challenging than their peers' students in terms of ability. The treatment-control differences were not statistically significant, which means we did not find evidence of strategic assignment of students.[48]

Figure IV.4. Assignment of More or Less Academically Challenging Students to Classrooms, Teacher Perceptions by Treatment and Focal Teacher Status



Source: 2009–2010 Mathematica Teacher Background Survey.

Note: N = 61 treatment focal, 41 control focal, 112 treatment non-focal, and 111 control non-focal respondents. Treatment-control differences are not statistically significant at the 0.05 level using a two-sided test.

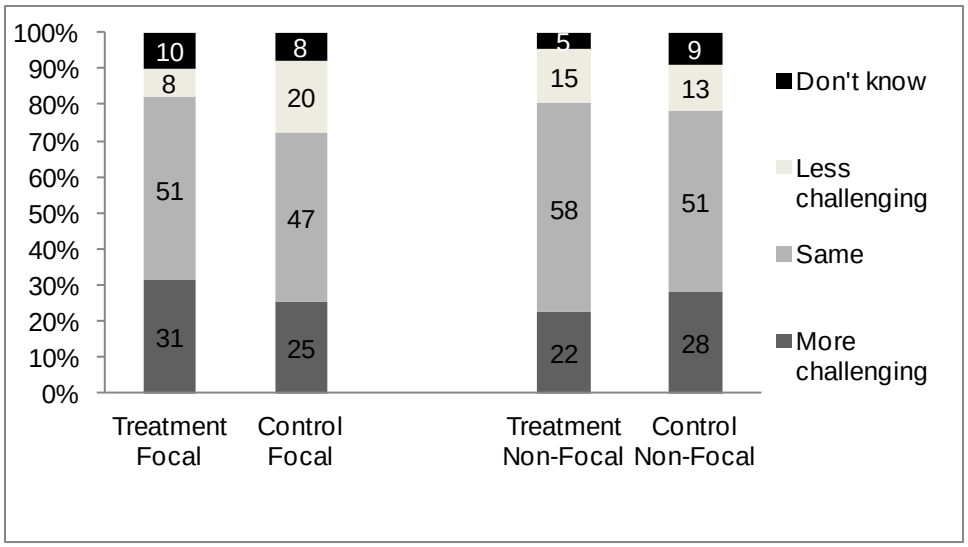
[47] The survey item was worded as follows: "Think about the ABILITY LEVELS of the students assigned to you class(es) this year compared to those of student assigned to your colleague(s) teaching the same grade level or subjects in your school. Would you say the students in YOUR class(es) are…

a. More challenging in ability
b. About the same level of ability
c. Less challenging in terms of ability
d. Cannot judge. I am unfamiliar with the ability levels of the students in the other class(es)."

[48] To check the credibility of the survey responses, we conducted a separate analysis in which we classified teacher teams by whether the responses within the teams were consistent or inconsistent with each other. The teams were coded as consistent if they had at least one teacher saying he or she had less challenging students for every team with at least one teacher reporting more challenging students, and vice versa. Twenty percent of treatment teams and 23 percent of control teams were inconsistent by this definition. Another 42 percent (or 64 percent if we use a stricter definition) of treatment and 40 percent (55 percent with the stricter definition) of control teams were ambiguous because there was at least one survey nonrespondent whose answer could possibly have offset the response of the team's teachers who did respond. This pattern suggests that subjective response items have some social desirability bias, but the bias was not likely different between teachers in treatment and control teams.

Figure IV.5 shows the corresponding percentages of the same groups of teachers for the question about how challenging the students were in terms of *behavior*.[49] As with the academically challenging students, teachers the most common response was for teachers to report having the "about the same" level of behavioral challenges as their peers rather than "more" or "less" challenging. However, they more frequently reported having "more" challenging than "less" challenging students in all four groups (31 versus 8 percent for treatment focal, 25 versus 20 percent for control focal, 22 versus 15 percent for treatment non-focal, and 28 versus 13 percent for control non-focal). The relationships between teacher perception of students' behavioral challenges and treatment status were not statistically significant for focal or non-focal teachers, meaning that the differences were not larger than we might observe by chance with samples of these particular sizes (102 focal and 223 non-focal teachers).

Figure IV.5. Assignment of More or Less Behaviorally Challenging Students to Classrooms, Teacher Perceptions by Treatment and Focal Status



Source: 2009–2010 Mathematica Teacher Background Survey.

Note: N = 61 treatment focal, 41 control focal, 112 treatment non-focal, and 111 control non-focal respondents. Treatment-control differences are not statistically significant at the 0.05 level using a two-sided test.

[49] The survey item was worded as follows: "Think about the DISCIPLINARY ISSUES of the students assigned to your class(es) this year compared to those of student assigned to your colleague(s) teaching the same grade level or subjects in your school. Would you say the students in YOUR class(es) are…

a. More challenging in terms of disciplinary issues
b. About the same in terms of disciplinary issues
c. Less challenging in terms of disciplinary issues
d. Cannot judge. I am unfamiliar with the disciplinary issues of the students in the other class(es)."

c. Principal Reports

We asked principals directly how students were assigned to classrooms in the specified grade teams.[50] The results, shown in Table IV.6, suggest that the most common assignment mechanism was random or balanced assignments, with principals reporting that students in 57 percent of control teams and 47 percent of treatment teams reportedly assigning students this way. The difference was not statistically significant (p-value = 0.294). Approximately one-quarter of treatment teams and 19 percent of control teams had their student rosters formed by matching students to teachers, according to principals. Ability grouping was reported by principals to be used in 12 and 13 percent of treatment and control teams, respectively. None of the treatment-control differences was statistically significant.

Table IV.6. How Students Were Assigned to Classrooms, Principal Report

Method of Assignment	Treatment Percentage	Control Percentage	Impact	P-value
Random (or similar method to balance academic level, gender, and/or behavioral problems)	47.4	57.4	-10.0	0.294
Matching Student Needs to Teachers' Specific Abilities	24.6	18.5	6.0	0.444
Homogeneous Groups Based on Ability or Course Difficulty	12.3	13.0	-0.7	0.915
Looping or Related Approach to Keep Previous Year Student Rosters Mostly Intact	7.0	3.7	3.3	0.445
Other	8.8	7.4	1.4	0.795

Source: 2009–2010 Mathematica Principal Survey.

Note: N = 57 treatment teams, 54 control teams. None of the treatment-control differences are statistically significant at the 0.05 level.

3. How Were Mentoring and Other Resources Allocated?

Strategic assignment of students is one way in which principals might adapt to the opportunity to fill vacancies with transfer teachers in a program like TTI. Another way to take advantage of the opportunity is to shift resources that would have gone toward supporting a new, presumably inexperienced teacher to other teachers instead, or perhaps to reduce the level of mentoring in that grade level altogether. In such cases, there could be more resources available to other grade-level teams in the same school.

[50] The survey item was worded as follows: "Which ONE of the following statements best describes how students were assigned to classrooms/teachers in [the given grade] for 2009-2010. Students were assigned:

a. At random (or similar method to ensure balance of academic level, gender, and/or behavioral problems).
b. By matching student needs to teachers' specific abilities.
c. By creating homogeneous groups based on ability or course difficulty.
d. By 'looping' or a related approach to keep previous year student rosters mostly intact.
e. Other (please specify)"

Figure IV.6 shows the level of support that teachers reported receiving, shown separately for treatment focal, control focal, treatment non-focal, and control non-focal teachers.[51] Treatment focal teachers (made up mostly of TTI transfers) did report having a mentor at a significantly lower rate than control focal teachers (39 versus 66 percent, p-value = 0.007). The differences in time per week on average, 33 versus 58 minutes, were not statistically significant (p-value = 0.105).[52] The non-focal teachers, both treatment and control, reported receiving levels of support that were in between the averages for treatment focal and control focal teachers, consistent with the hypothesis that typical teachers require more support than a TTI transfer but less support than someone who is new to the school or grade. None of the differences between treatment and control non-focal teachers was statistically significant.

4. Were TTI Teachers Used in Mentoring or Leadership Roles or Given Other Duties?

Yet another way to take advantage of TTI might be to assign additional duties or responsibilities to transfer teachers. The design of the intervention did not require principals to create or require any special duties or roles as a condition of being hired or receiving the TTI bonus, but there was no restriction against a principal imposing such a condition or simply assigning the teacher or requesting that the teacher fill such a role.

Table IV.7 shows the percentages of teachers by treatment and focal status who reported playing such roles. The evidence suggests that treatment focal teachers provided more mentoring to their peers than did control focal teachers (25 minutes versus less than 1 minute per week). Of the non-focal teachers, 18 and 20 percent of treatment and control group teachers reported providing mentoring and the amounts were just over half an hour per week for both groups. The differences for non-focal teachers were not statistically significant and therefore there was not strong evidence that the higher level of mentoring provided by the treatment focal teachers resulted in an offsetting decrease in that provided by their peers.

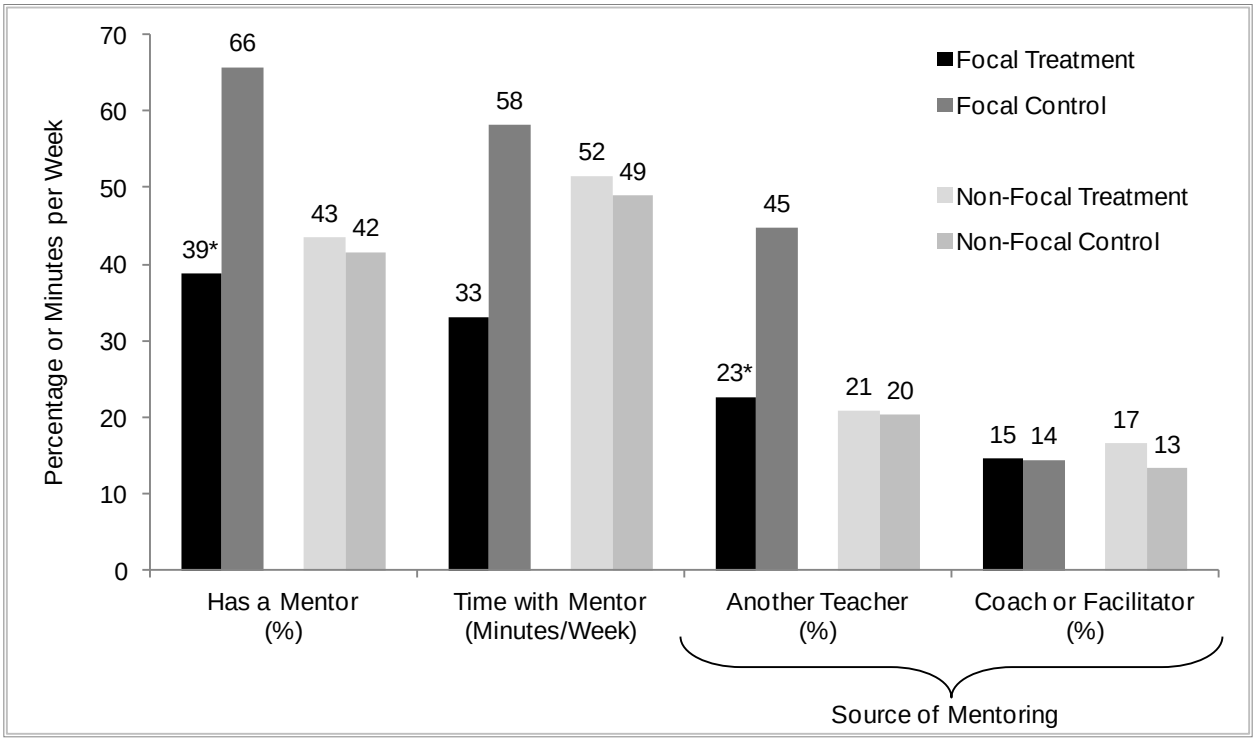
[51] The survey item was worded as follows: "To what extent are teachers who are new to this school during 2009-2010 provided with the following types of support?

a. Mentoring
b. Routine visits from principal
c. Routine visits from Curriculum specialist of veteran teacher
d. Oversight by district administration
e. Other"

Principals were asked to respond for novice teachers, new to school teachers, and new to district teachers.

[52] Given the observed variation in time spent with mentors and the size of the sample (103 focal teachers), the study was only able to detect an impact on time spent with a mentor if it had been at least 30 minutes per week. A future report will draw upon a larger sample, when all ten study districts are included.

Figure IV.6. Mentoring Received by Teachers, by Treatment and Focal Teacher Status



Source: 2009–2010 Mathematica Teacher Background Survey.

Note: N = 62 treatment focal, 41 control focal, 115 treatment non-focal, and 113 control non-focal teachers.

* Treatment-control difference (within focal status) is statistically significant at the 0.05 level using a two-sided test.

None of the treatment-control differences in the rates at which focal or non-focal teachers played leadership roles was statistically significant. The survey asked teachers, "In which of the following activities are you currently involved at your school? (a) Serving as a grade level or subject area chair. (b) Serving on a school improvement committee. (c) Working to obtain external funding for my school (i.e., grants or funding from external organizations for projects/supplies/materials. (d) Leading or promoting teacher collaboration. (e) Observing or providing feedback to other teachers. (e) Other (Please specify)." Results are shown in the bottom of Table IV.7.

Table IV.7. Involvement in Mentoring and School Leadership, by Focal and Non-Focal Teachers

	Focal Teachers			Non-Focal Teachers		
Outcome[a]	Treatment	Control	Difference	Treatment	Control	Difference
Support Provided						
Number of teachers mentored	0.4	0.0	0.3	0.6	0.6	0.1
Average minutes per week of mentoring provided	24.9	0.4	24.5*	34.2	31.8	2.4
Leadership Roles						
Serves as grade-level or subject area chair	16.1	13.2	3.0	32.2	40.7	-8.5
Serves on school improvement committee	25.8	34.2	-8.4	43.5	45.5	-2.1
Works to obtain external funding for the school	17.7	14.5	3.3	13.9	15.9	-2.0
Leads or promotes teacher collaboration	54.8	47.4	7.5	47.8	52.7	-4.9
Observes or provides feedback to other teachers	29.0	31.1	-2.0	43.9	46.4	-2.5
Involved in other activities	33.9	33.3	0.6	38.8	26.7	12.0
Sample Size[b]	**62**	**41**		**115**	**113**	

Source: 2009–2010 Mathematica Teacher Background Survey.

[a] Units are percentages unless otherwise indicated.

[b] Sample size is number of teachers.

* Differences are statistically significant at the 0.05 level using a two-sided test.

This page has been left blank for double-sided copying.

V. SUMMARY AND NEXT STEPS

This report described the implementation of a transfer incentive program in seven large and diverse school districts. Districts and schools were recruited, the highest-performing teachers were identified, and 63 teachers made the decision to move to a low-achieving school, filling 90 percent of the vacancies targeted by the program.

The impacts of the transfer incentive program on student and teacher outcomes will be examined in a subsequent report. This chapter briefly summarizes some of the implementation experiences and early (intermediate) impacts on teacher hiring, teacher support, and student assignment and practices. As this is an interim report, we discuss the next steps of the evaluation.

A. Implementation

We posed four implementation questions and addressed them with a combination of survey data, district administrative records, and program data.

How were candidates identified and recruited? We documented the effort that was required and the rules used to identify each district's highest-performing teachers using value-added analysis and to identify those who were eligible for the transfer incentive. We also documented the efforts of site managers, who played the critical role of hosting information sessions for teachers and principals, maintaining regular communication with teachers and principals, and helping both sides make matches by arranging interviews and encouraging their continued interest in the goals of TTI. Identifying teachers required a team of data specialists and programmers to spend one to three months (between January and March) depending on the district. Recruiting principals and transfer candidates took a site manager working approximately half-time per district about four months (April through August) to complete all of the transfers.

How responsive were teachers to the transfer incentive? Specifically, how many of the districts' highest-performing teachers went through the entire process of applying, interviewing, and ultimately transferring to a low-achieving school? In the seven districts, just over 1,000 teachers were identified as highest performing in their district and eligible to participate in the transfer program. Of these candidates, 38 percent attended an information session, 24 percent applied for a transfer position, and 13 percent (135 individuals) were interviewed by a principal. All told, among the 70 vacancies assigned to the treatment condition, 63 positions (90 percent of the targeted vacancies) were filled with a TTI candidate by the end of recruitment season, shifting six percent of the highest-performing teachers to the targeted schools. These numbers suggest that a large initial pool of candidates may be required to yield the desired number of successful transfers.

What types of teachers applied and transferred? Relative to the teachers who did not apply, teachers who applied to TTI were different in some consistent ways. They were more likely to be African American, be unmarried, have a lower salary, or be less satisfied with their current school policy. In the four districts where we estimated value-added scores and had student-level data, teachers with a higher percentage of disadvantaged current students and teachers who were in the top 10 percent of value-added ranking—the better among the highest-performing teachers—were more likely to apply.[53] Among those teachers who applied, teachers who went through the entire process and transferred to a low-achieving school were not significantly different in terms of any background characteristics than those who did not transfer. That is, once the decision to apply to the incentive program was made, none of the personal background characteristics defined who decided to transfer. However, transfer teachers were more likely than other characteristics to be satisfied with their current students.

Where did these teachers transfer from? Specifically, were they coming from schools that are similar to the ones into which they are transferring, and teaching students who are similar in both schools, or were there differences between the schools and students in key characteristics? In terms of school rank in student achievement, the average sending school was ranked in the 60th percentile (where the 100th percentile is the highest-scoring school in the district). The average receiving school was ranked in the 18th percentile. In terms of poverty rank, based on percentage eligible for FRL, the average sending school was in the 55th percentile and the average receiving school was in the 18th percentile.

We also compared the students taught by transfer teachers in their original school and the school to which they transferred. The average transfer teacher for whom data were available taught students in his or her new school who were significantly less likely to be white and more likely to be African American or Hispanic and low income (measured by FRL status) than those in the school from which he or she transferred.

B. Intermediate Impacts

As a precursor to examining the impacts on student achievement and teacher retention in the next report, we examined here how the transfer incentive affected the school itself. These are intermediate impact questions.

Who filled the vacancies? First we looked at the types of teachers hired into the previously vacant positions. The control teams filled their vacancies in a variety of ways, including new hires, but also by accepting transfers from elsewhere in the district, moving staff from elsewhere in the school, and sometimes undoing what was a vacancy by hiring back a laid off teacher or convincing an outgoing teacher to stay. On the treatment side, we showed that 90 percent of the vacancies were filled by TTI candidates. The remaining slots were either filled by someone outside the program or the vacancy was lost. The types of teachers in treatment teams were significantly more experienced, with a significant difference of five years in the classroom, based on our analysis of focal teachers.

[53] We are not able to characterize the extent to which this was true in the three districts for which we had insufficient data.

Did the transfer incentive affect school climate? Data from principal reports did not support the claim that providing $20,000 payments to teachers with high value-added scores would undermine collegiality or harm collaboration and sharing of ideas within the teacher team. Treatment-control differences in the principal reports of various school climate measures were not statistically significant.

Did principals assign students differently as a result of the transfer incentive? We did not find evidence of this behavior based on principal self-reports. We hypothesized that schools might strategically assign students in response to a new type of teacher. We examined the data on student background and asked teachers and principals directly through surveys about how students were assigned to classrooms. We did not find statistically significant impacts from any of these sources suggesting that students were assigned differently as a result of the transfer incentive.

Were mentoring and leadership roles assigned differently? Another hypothesized effect would be to allow schools to shift mentoring resources and other staff time to take advantage of a presumably more experienced, accomplished teacher who enters via the transfer incentive than the teacher who would have filled the given vacancy. We did find that focal teachers in treatment teams versus control teams were less likely to have a mentor and less likely to be mentored by a fellow teacher. They were also more likely to provide mentoring and spent more time than control focal teachers providing such assistance to others. The differences were statistically significant.

C. Next Steps

The next step in the study is to estimate the impacts of the transfer incentives on student achievement test scores and the retention of highest-performing teachers. The analyses in this report, especially in Chapter IV, are critical for understanding the potential for indirect and resource allocation effects. These provide the context for how we interpret the impact findings and whether we pay closer attention to the *team-level impacts*, which include an average of direct and indirect impacts, or the *focal teacher-level impacts*.

This page has been left blank for double-sided copying.

REFERENCES

Buddin, Richard, and Gema Zamarro. "Teacher Quality, Teacher Licensure Tests, and Student Achievement." Santa Monica, CA: RAND, May 2008.

Carroll, Stephen, Robert Reichardt, and Cassandra Guarino. "The Distribution of Teachers Among California's Districts and Schools." Santa Monica, CA: RAND, October 2000.

Clotfelter, Charles, Helen Ladd, Jacob Vigdor, and Justin Wheeler. "High Poverty Schools and the Distribution of Teachers and Principals." Washington, DC: National Center for Analysis of Longitudinal Data in Education Research, December 2006.

Education Trust. "Their FAIR Share: How Texas-Sized Gaps in Teacher Quality Shortchange Low-Income and Minority Students." Washington, DC: EdTrust, February 2008.

Engel, Mimi and Brian Jacob. "New Evidence on Teacher Labor Supply." Cambridge, MA: National Bureau of Economic Research, Working Paper No. 16802. February 2011.

Glazerman, Steven, Daniel Mayer, and Paul Decker. (2006) "Alternative Routes to Teaching: The Impacts of Teach For America on Student Achievement and Other Outcomes." *Journal of Policy Analysis and Management*, vol. 25, no. 1. pp. 75-96.

Glazerman, Steven, and Allison Seifullah. "An Evaluation of the Teacher Advancement Program (TAP) in Chicago: Year Two Impact Report." Washington, DC: Mathematica Policy Research, May 2010.

Glazerman, Steven, Eric Isenberg, Sarah Dolfin, Martha Bleeker, Amy Johnson, Mary Grider, and Matthew Jacobus. "Impacts of Comprehensive Teacher Induction: Final Results from a Randomized Controlled Study." Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, June 2010.

Glazerman, Steven, and Jeffrey Max. "Do Low-Income Students Have Equal Access to the Highest-Performing Teachers?" *NCEE Evaluation Brief*. Washington, DC: U.S. Department of Education, Institute of Education Sciences. April 2011.

Goldhaber, Dan. "Addressing the Teacher Qualifications Gap: Exploring the Use and Efficacy of Incentives to Reward Teachers for Tough Assignments." Washington, DC: Center for American Progress, November 2008.

Gordon, Robert, Thomas Kane, and Douglas Staiger. "Identifying Effective Teachers Using Performance on the Job." Washington, DC: Hamilton Project, Brookings Institution, April 2006.

Huber, Peter J. "The Behavior of Maximum Likelihood Estimation under Nonstandard Conditions." *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability 1*, edited by L.M. LeCam and J. Neyman. Berkeley, CA: University of California Press, 1967.

Isenberg, Eric, and Heinrich Hock. "Design of Value-Added Models for IMPACT and TEAM in DC Public Schools, 2010-2011 School Year." Washington, DC: Mathematica Policy Research, May 2011.

Jackson, C. Kirabo, and Elias Brugmann. "Teaching Students and Teaching Each Other: The Importance of Peer Learning for Teachers." Cambridge, MA: National Bureau of Economic Research, Working Paper No. 15202. July 2009.

Jordan, Heather, Robert Mendro, and Dash Weerasinghe. "Teacher Effects on Longitudinal Student Achievement." Paper presented at the CREATE annual meeting, Indianapolis, IN, July 1997.

Lankford, Hamilton, Susanna Loeb, and James H. Wyckoff. "Teacher Sorting and the Plight of Urban Schools: A Descriptive Analysis." *Educational Evaluation and Policy Analysis*, vol. 24, no. 1, 2002, pp. 37–62.

Lipscomb, Stephen, Bing-ru Teh, Brian Gill, Hanley Chiang, and Antoniya Owens. "Teacher and Principal Value-Added: Research Findings and Implementation Practices." Cambridge, MA: Mathematica Policy Research, September 2010.

McCaffrey, Dan F., J. R. Lockwood, Daniel Koretz, Thomas A. Louis, and Laura Hamilton, "Models for Value-Added Modeling of Teacher Effects," *Journal of Educational and Behavioral Statistics*, vol. 29, no. 1, 2004, pp. 67-101.

Morris, Carl N. "Parametric Empirical Bayes Inference: Theory and Applications." *Journal of American Statistical Association*, vol. 78, no. 381, 1983, pp. 47-55.

Peske, Heather, and Kati Haycock. "Teaching Inequality: How Poor and Minority Students Are Shortchanged on Teacher Quality." Washington, DC: The Education Trust, June 2006.

Presley, Jennifer, Bradford White, and Yuqin Gong. "Examining the Distribution and Impact of Teacher Quality in Illinois." Edwardsville, IL: Illinois Education Research Council, 2005.

Rivkin, Steven, Erick Hanushek, and John Kain. "Teachers, Schools, and Academic Achievement." *Econometrica*, vol. 73, no. 2, March 2005, pp. 417–458.

Rockoff, Jonah, Brian Jacob, Thomas Kane, and Douglas Staiger. "Can You Recognize an Effective Teacher When You Recruit One?" Cambridge, MA: National Bureau of Economic Research, November 2008.

Sanders, William, and June Rivers. "Cumulative and Residual Effects of Teachers on Future Student Academic Achievement." Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center, November 1996.

Sass, Tim, Jane Hannaway, Zeyu Xu, David Figlio, and Li Feng. "Value Added of Teachers in High-Poverty Schools and Lower-Poverty Schools." CALDER Working Paper No. 52. Washington, DC: Urban Institute, November 2010.

Springer, Matthew, Dale Ballou, Laura Hamilton, Vi-Nhuan Le, J.R. Lockwood, Daniel McCaffrey, Matthew Pepper, and Brian Stecher. "Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching." Nashville, TN: National Center on Performance Incentives, Vanderbilt University, 2010.

Tennessee Department of Education. "Tennessee's Most Effective Teachers: Are They Assigned to the Schools That Need Them Most?" Nashville, TN: Tennessee Department of Education, 2007.

White, Halbert. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica*, vol. 48, no. 4, 1980, pp. 817-830.

APPENDIX A

ANALYSIS OF SURVEY NONRESPONSE

This page has been left blank for double-sided copying.

Table A.1. Respondents Versus Full Sample of Respondents and Non-respondents (percentages)

	Candidate Survey	
Subgroup	Respondents	Full Sample
District		
A	17.8	18.0
B	8.9	8.2
C	31.5	32.9
D	7.5	7.1
E	9.0	8.9
F	13.9	13.2
G	11.4	11.7
Total	100.0	100.0
Pool		
Elementary	66.2	65.3
Middle school English language arts	18.1	19.0
Middle school math	15.7	15.7
Total	100.0	100.0
In top 10% of value-added distribution		
Elementary pool	48.7	50.2
Middle school English language arts pool	52.9	54.2
Middle school math pool	56.2	52.2
Application status		
Did not apply	72.1	75.6*
Applied but did not transfer	20.4	18.2*
Transferred	7.5	6.2*
Total	100.0	100.0
Sample size[†]	**835**	**1,012**

* Difference in candidate survey respondents versus full sample distributions for application statues are statistically significant at the 0.05 level using chi-square test of independence.

[†] Sample size for comparing whether candidates are in the top 10 percent of value-added ranking in their pool between respondent and the full sample is different because we only have value-added ranking for candidates in the 4 districts where we conducted value-added analysis ourselves.

Table A.2. Survey Completion Rates by Subgroup, Teacher and Principal Surveys (percentages)

	Teacher Background Survey			Principal Survey (One response per team)		
Subgroup	Treatment	Control	Difference	Treatment	Control	Difference
All	76	75	0	94	90	4
District						
A	85	72	12	100	100	0
B	89	87	3	83	100	-17
C	63	50	13	100	67	33
D	65	95	-30*	86	100	-14
E	91	63	28*	83	100	-17
F	69	76	-8	100	82	18*
G	72	79	-7	100	80	20
Grade						
3	74	81	-7	85	100	-15
4	83	78	5	100	100	0
5	77	77	0	100	89	11
6	92	77	15	83	75	8
7	54	46	8	100	80	20
8	59	59	0	83	71	12
School Poverty						
Lower poverty (<=80% FRPL)	79	82	-3	90	89	1
Higher poverty (>80% FRPL)	74	72	1	95	90	5
School Race/Ethnicity						
Majority African American	76	80	-4	92	91	1
Majority Hispanic	73	76	-3	97	100	-3
Majority white	44	29	16	100	50	50
No Majority	93	60	33*	88	67	21
School Size						
Smaller (<=700 students)	79	72	7	93	90	3
Larger (>700 students)	71	83	-12	95	92	4
Sample Size	**241**	**209**		**63**	**60**	

* Difference is statistically significant at the 0.05 level using a two-sided test.

Table A.3. Respondents Versus Full Sample of Respondents and Non-respondents (percentages)

	Teacher Background Survey		Principal Survey (One response per team)	
Subgroup	Respondents	Full Sample	Respondents	Full Sample
District				
A	17.4	16.7	15.9	14.6
B	8.9	7.6	8.9	8.9
C	5.0	6.4	5.3	5.7
D	10.0	9.6	9.7	9.8
E	14.5	14.0	18.6	18.7
F	35.7	37.1	33.6	34.2
G	8.6	8.7	8.0	8.1
Total	100.0	100.0	100.0	100.0
Grade				
3	29.5	28.9*	22.1	22.0
4	27.7	25.8	25.7	23.6
5	25.7	25.1	28.3	27.6
6	6.5	5.8	7.1	8.1
7	3.8	5.8	8.0	8.1
8	6.8	8.7	8.9	10.6
Total	100.0	100.0	100.0	100.0
School Poverty				
Lower poverty (<=80% FRPL)	34.2	32.2	31.0	31.7
Higher poverty (>80% FRPL)	65.8	67.8	69.0	68.3
Total	100.0	100.0	100.0	100.0
School Race/Ethnicity				
Majority African American	46.0	44.4*	45.1	45.5
Majority Hispanic	41.9	42.4	42.5	39.8
Majority white	1.8	3.6	2.7	3.3
No Majority	10.3	9.6	9.7	11.4
Total	100.0	100.0	100.0	100.0
School Size				
Smaller (<=700 students)	64.3	64.4	72.6	73.2
Larger (>700 students)	35.7	35.6	27.4	26.8
Total	100.0	100.0	100.0	100.0
Sample Size	**339**	**450**	**113**	**123**

* Difference in teacher survey respondents versus full sample distributions for grade and school race/ethnicity are statistically significant at the 0.05 level using chi-square test of independence.

APPENDIX B

VALUE-ADDED ANALYSIS TO IDENTIFY HIGHEST-PERFORMING TEACHERS

This page has been left blank for double-sided copying.

The first step for the Talent Transfer Initiative (TTI) was to identify the highest-performing teachers in each study district. To do this, we estimated teachers' value added to student achievement based on two or more years of test score data from state assessments. Value added represents the amount of learning growth that can be attributed to the teacher, holding constant the factors outside the teacher's control. It can be estimated by measuring student achievement growth over time and comparing the actual scores of each teacher's students to the predicted scores given the prior achievement and possibly other characteristics of that teacher's students. It requires longitudinal data and a reliable student-teacher link. Using more than one year of data is meant to increase the statistical precision and stability of the estimates, identifying teachers with high persistent performance—in other words, a strong track record.

The value-added estimates were prepared by either the participating districts themselves, working with an outside vendor, or the study team. One of the study districts had its vendor conduct the data analysis, and supplied the TTI team with a list of teachers the district identified as being highest performing based on these pre-existing measures of teacher effectiveness. Two other districts supplied the TTI team with information from their vendor on teachers' value-added estimates, which the study team then combined across years and used to identify the top performers.[54] For the remaining four districts, Mathematica used raw data on student achievement, demographics, and enrollment to link students to teachers and then computed teachers' value added. The approach used by Mathematica for its four districts is described below, but the other three districts followed a similar approach in their estimation of teacher effectiveness. Mathematica did not attempt to duplicate the methods used by the other districts. Instead, the goal was to estimate a model that could plausibly have been adopted by the district in regular implementation of an intervention such as TTI.

Estimation Equation

We estimated a value-added model separately for three pools of teachers: elementary school teachers, middle school math teachers, and middle school English language arts (ELA) teachers. Elementary school included grades three to five or four to five, and middle school included grades six to eight. We used three waves of student achievement growth data (between 2004-05 and 2005–06, between 2005–06 and 2006–07, and between 2006–07 and 2007–08) to identify highest-performing teachers.

All of a teacher's student observations for a particular year were dropped from the estimation sample if the teacher was linked to fewer than five students' test scores in that year. Students who spent less than 20 percent of the school year with a teacher were also excluded from the estimation sample for that teacher.

The estimation equation is:

$$(1)\ Y_{ijt} = \lambda_{t-1} * Y_{ij,t-1} + \alpha_1 * X_{ijt} + \alpha_2 * Z_{jt} + \beta_j * D_{ijt} + e_{ijt}$$

[54] The same vendor, the SAS Institute, conducted the value-added analysis for each of the three districts. The methods used by the SAS Institute are described here: http://www.sas.com/govedu/edu/k12/evaas/index.html.

where Y_{ijt} is the post-test score for student i who is taught by teacher j in year t; $Y_{ij,t-1}$ is the pre-test score for that same student, which is assumed to capture previous inputs into student achievement; and e_{ijt} is the error term. X_{ijt} is a vector of control variables that includes the following student-level variables: indicators for gender, race/ethnicity, free or reduced-price lunch (FRL) status,[55] English language learner status, special education status, disability type, grade repetition status, and overage for grade status.[56] Z_{jt} includes the following teacher-level variables: the percentage of a teacher's students who were mobile, the percentage of a teacher's students who were grade repeaters, and class size. Grade-by-year dummies are also included in Z_{jt} to eliminate any mean differences between grade levels and years. Dosage (D_{ijt}) is a variable that equals the percentage of the year student i in year t was taught by teacher j, and 0 if student i was not taught by teacher j in year t. D_{ijt} is expressed as a vector of such dosage variables that includes separate values for each teacher-year. The coefficients λ_{t-1}, α_1, α_2, and β_j are parameters to be estimated. The performance measures ("teacher effects") are contained in the vector β_j, which is the set of coefficients of the dosage variables.

After initial estimation of the teacher effects, we standardized subject-specific performance measures (one for math and one for language arts, if applicable) within each grade level.[57] We then excluded from the rankings any teachers who had fewer than two years of subject-specific performance measures. While some elementary schools are departmentalized, the majority of elementary school teachers taught in self-contained classrooms. For these teachers, performance measures were calculated by taking the average of their math and ELA performance measures. The top 20 to 25 percent of teachers in each of the three pools—elementary school teachers, middle school math teachers, and middle school ELA teachers—were identified as being the highest-performing teachers in their respective districts.

Controlling for Measurement Error in the Pre-Test

Prior to estimating Equation (1), we correct for measurement error in the pre-test by fitting an errors-in-variables regression model.[58] We obtain the reliability for each test, when available, from either the test publisher or the school district. We employ a two-stage procedure. In the first stage, we estimate the following errors-in-variables regression model by using the average published reliability of the test across grades and years to remove the bias caused by the measurement error in the pre-test:[59]

$$(2)\ Y_{ijt} = \lambda_{t-1} * Y_{ij,t-1} + \alpha_1 * X_{ijt} + \beta_j * D_{ijt} + e_{ijt}$$

[55] One district did not provide data on FRL

[56] Missing values in Y_{ijt}, $Y_{ij,t-1}$ and X_{ijt} were imputed with predicted values from a regression model.

[57] This assumes that the distribution of teacher effectiveness is the same in each grade within a district, but has the benefit of removing any artificial differences associated, for example, with the properties of the assessment instrument and the ways such properties vary by grade.

[58] We implement this model by using the eivreg command in Stata.

[59] The errors-in-variables correction works by subtracting the reliability from the diagonal terms of the regression crossproduct matrix. The resulting parameters are consistent for the normal distribution. See Isenberg and Hock (2011) for a recent application.

The control variables for student background characteristics in Equation (2) are identical to those used in (1). Using $\hat{\lambda}_{t-1}$, the estimated value for the coefficient of the pre-test from Equation (2), we calculate the estimated adjusted gain for each student in each year:

(3) $\hat{G}_{ijt} = Y_{ijt} - \hat{\lambda}_{t-1} * Y_{ij,t-1}$

The second-stage regression model pools the data from all years and uses the adjusted gain as the dependent variable:

(4) $\hat{G}_{ijt} = \alpha_1 * X_{ijt} + \alpha_2 * Z_{jt} + \beta_j * D_{ijt} + e_{ijt}$

In Equation (4), we account for the correlation in outcomes for students in different years by using robust standard errors (Huber 1967; White 1980). This errors-in-variables measurement error correction method underestimates the standard errors of β_j because it treats $\hat{\lambda}_{t-1}$ as identical to its true value, λ_{t-1}; if $\hat{\lambda}_{t-1}$ is estimated precisely, it will be negligible. By substituting Equation (3) into (4), rearranging terms, and treating $\hat{\lambda}_{t-1}$ as λ_{t-1}, we arrive at Equation (1).

Shrinkage Estimator

After estimating Equation (1) to obtain performance measures from the β_j coefficients, we apply a shrinkage procedure outlined in Morris (1983) to calculate empirical Bayes performance measures and standard errors. Using this procedure, the empirical Bayes estimate of each performance measure is approximately the precision-weighted average of the original performance measure (an individual element of the β_j vector) and the mean of all the point estimates (all the elements of β_j):

(5) $$\beta_j^{EB} \approx \left(\frac{\frac{1}{\sigma_j^2}}{\frac{1}{\sigma_j^2} + \frac{1}{\sigma_\beta^2}} \right) \beta_j + \left(\frac{\frac{1}{\sigma_\beta^2}}{\frac{1}{\sigma_j^2} + \frac{1}{\sigma_\beta^2}} \right) \mu_\beta ,$$

where β_j^{EB} is the empirical Bayes estimate of an element of the β_j vector, β_j is the original point estimate, σ_j is the standard error of the original point estimate, μ_β is the mean of all the point estimates, and σ_β is the standard deviation of all the point estimates.

Due to the precision weighting of the original estimate and the mean of all the point estimates, the empirical Bayes performance measure is designed to place relatively more weight on the mean when the original estimate has a high standard error. This is especially important for a program like TTI because the focus is on the upper tail of the teacher performance distribution. Random estimation error will vary across teachers when we try to estimate their value added, because they have different numbers of students, their students can be more or less homogeneous, and their students' characteristics can be more or less similar to the population average. Each of these factors influences the precision of the individual teacher's value-added

estimate. Most importantly, if that precision does vary, the most imprecisely estimated teacher effects will be overrepresented in both tails of the distribution (because the variance in the effect estimates will contain both true variation in teacher quality plus a larger error variance). As a result, a program like TTI would identify an artificially high number of teachers with small classes or outlier students unless the estimates were corrected. The empirical Bayes shrinkage adjusts the estimates to account for this phenomenon.

Diagnostics

We also conducted a series of robustness checks to ensure the stability of the rankings generated for the model described above: excluding all control variables except for year and grade dummies, estimating the model without controlling for measurement error in the pre-test, including higher-order terms of the pre-test variables, and estimating the model separately by each of the three school years.

APPENDIX C

SUPPLEMENTAL TABLES AND FIGURES

This page has been left blank for double-sided copying.

Table C.1. Candidate Interview Process and Perceptions by Transfer Status (percentages)

Perception of Interview		Interviewed but Did Not Transfer	Transferred
Interview was informative		74.3	91.2
Had opportunities to communicate strengths during the interview		91.1	94.1
Principal/interviewer was genuinely interested		71.3	86.8
Principal/interviewer responded to candidate's questions about the position		90.0	91.5
Principal seemed like someone the candidate could work with		73.0	84.8
Interview increased candidate's desire to teach at the school		48.0	71.2
Sample Size	**Number of Interviews**	**101**	**102**
	Number of Candidates	**72**	**63**

Source: 2009–2010 Mathematica TTI Candidate Survey

Table C.2. Candidate Interview Structure by Transfer Status (percentages)

Interview Structure		Interviewed but Did Not Transfer	Transferred
Had a one-on-one interview with the principal or assistant principal		63.4	57.1
Interviewed with other school staff		58.4	66.0
Asked to give a teaching demonstration		10.9	10.5
Given a tour of the school		27.7	48.1
Met students at the school		4.0	12.4
Sample Size	**Number of Interviews**	**101**	**102**
	Number of Candidates	**72**	**63**

Source: 2009–2010 Mathematica TTI Candidate Survey

Table C.3. Factors Related to the Probability of Applying

	In All Seven Districts		In Four Districts With Student Data [b]	
Factor[a]	Odds Ratio	Standard Error	Odds Ratio	Standard Error
Dependent Variable: Probability of Applying to TTI				
Income (dollars)	1.00*	(0.00)	1.00	(0.00)
Demographic Variables				
Male	1.07	(0.31)	1.59	(0.90)
African American	2.49*	(0.64)	1.92	(0.95)
Hispanic	1.42	(0.46)	0.42	(0.26)
Married	0.51*	(0.10)	0.45*	(0.15)
Married with co-residing children under 5	1.63	(0.44)	1.13	(0.50)
Residential Variables				
Owns home	0.73	(0.21)	0.48	(0.23)
Travel time in 2008-09 (minutes)	1.00	(0.01)	1.00	(0.02)
Professional Variables				
Has master's or above degree	1.27	(0.24)	1.77	(0.57)
Candidate for certification or certified	1.37	(0.31)	0.52	(0.24)
Satisfaction Indicators				
Satisfied with school policy	0.40*	(0.09)	0.30*	(0.12)
Satisfied with salary	1.50*	(0.29)	1.23	(0.43)
Satisfied with professional environment	0.66	(0.19)	0.76	(0.41)
Satisfied with facilities	1.22	(0.32)	1.79	(0.76)
Satisfied with students	1.20	(0.26)	1.84	(0.72)
District Indicators				
B	2.99*	(1.20)		
C	1.53	(0.55)		
D	2.38*	(0.99)		
E	1.04	(0.42)	0.19*	(0.10)
F	4.22*	(1.50)	2.15	(0.87)
G	2.24*	(0.88)		
In Top 10% of Value-Added Ranking			2.27*	(0.70)
Percent of FRL Current Students			1.04*	(0.01)
Constant	2.30	(1.49)	0.48	(0.53)
Sample Size	**767**		**312**	
Log-Likelihood	**-404.13**		**-138.36**	
Likelihood Ratio (LR) Chi-squared	**89.31**		**82.39**	
p-value of LR Chi-squared	**0.00**		**0.00**	

[a] Factors included are dummy variables unless otherwise noted.

[b] The four districts are those where we estimated value added and had student-level data.

* Coefficient is statistically significant at the 0.05 level, two-sided test.

Table C.4. Factors Related to the Probability of Transferring, Conditional on Applying

	In All Seven Districts		In Four Districts With Student Data [b]	
Factor [a]	Odds Ratio	Standard Error	Odds Ratio	Standard Error
Dependent Variable: Probability of Transferring to a Low-Achieving School				
Income (dollars)	1.00	(0.00)	1.00	(0.00)
Demographic Variables				
Male	0.79	(0.37)	0.40	(0.33)
African American	2.41	(1.16)	2.42	(2.03)
Hispanic	2.45	(1.35)	7.11*	(6.91)
Married	0.84	(0.35)	0.45	(0.41)
Married with co-residing children under 5	1.73	(0.91)	0.57	(0.55)
Residential Variables				
Owns home	1.20	(0.63)	2.98	(2.13)
Travel time in 2008-09 (minutes)	1.01	(0.01)	1.03	(0.03)
Professional Variables				
Has master's or above degree	0.77	(0.29)	1.89	(1.63)
Candidate for certification or certified	0.94	(0.42)	0.94	(0.97)
Satisfaction Indicators				
Satisfied with school policy	0.76	(0.30)	0.34	(0.23)
Satisfied with salary	1.75	(0.68)	1.44	(0.91)
Satisfied with professional environment	0.90	(0.39)	1.02	(0.78)
Satisfied with facilities	0.65	(0.29)	4.22	(3.53)
Satisfied with students	3.01*	(1.42)	4.98	(4.12)
District Indicators				
B	0.69	(0.51)		
C	0.30	(0.19)		
D	0.39	(0.37)		
E	0.43	(0.35)	0.94	(1.14)
F	0.70	(0.53)	0.49	(0.46)
G	0.51	(0.41)		
In Top 10% of Value-Added Ranking			0.37	(0.24)
Percent of FRL Current Students			0.99	(0.02)
Constant	0.19	(0.26)	0.06	(0.14)
Sample Size	**216**		**89**	
Log-Likelihood	**-114.19**		**-43.79**	
Likelihood Ratio (LR) Chi-squared	**23.84**		**31.04**	
p-value of LR Chi-squared	**0.30**		**0.04**	

[a] Factors included are dummy variables unless otherwise noted.

[b] The four districts are those where we estimated value added and had student-level data.

* Coefficient is statistically significant at the 0.05 level, two-sided test.

Table C.5. Characteristics of Teachers Who Filled Treatment and Control Vacancies, Using Inclusive Definition of Focal Control Teachers (Percentages)

Characteristic	All Treatment Focal	All Control Focal [a]	Treatment-Control Difference	P-value
Professional Background				
Years of Experience in Teaching (average years)	12.8	9.2	3.6*	0.014
Years of Experience in Teaching (percentages by category)				
1 (first year teaching)	0.0	16.6	-16.6*	0.002
2-5 years	11.8	34.8	-22.9*	0.003
6-10 years	37.7	17.9	19.8*	0.014
11+ years	50.5	30.8	19.7*	0.025
Has a Master's or Doctorate Degree	47.4	26.1	21.3*	0.014
Has National Board Certification	22.7	11.0	11.7	0.087
Transferred via TTI	95.0	0.0	95.0*	0.000
Personal Background				
Female	88.7	82.3	6.4	0.336
Race/Ethnicity				
White, non-Hispanic	45.9	50.7	-4.8	0.603
African American, non-Hispanic	31.1	31.6	-0.5	0.950
Hispanic or Latino	16.2	14.5	1.7	0.802
Age (years)	42.2	38.2	4.0	0.051
Married or Living with a Partner	59.9	62.1	-2.1	0.813
Homeowner	81.9	54.3	27.6*	0.001
Sample Size (number of teachers)	**63**	**67**		

Source: 2009–2010 Mathematica Teacher Background Survey.

[a] The control focal teachers are those identified using the "inclusive" definition. See Table IV.4 for results based on selective definition.

* Difference between treatment focal mean and control focal mean is statistically significant at the 0.05 level using a two-sided test.

Table C.6. Involvement in Mentoring and School Leadership, by Focal and Non-Focal Teachers Based on Inclusive Definition of Focal Teachers

	Focal Teachers			Non-Focal Teachers		
Outcome[a]	Treatment	Control	Difference	Treatment	Control	Difference
Support Provided						
Provides mentor support	14.6	6.7	7.8	18.3	20.4	-2.1
Number of teachers mentored	0.4	0.4	-0.0	0.6	0.3	0.3
Average minutes per week of mentoring provided	24.6	7.7	16.9	35.1	36.0	-0.9
Leadership Roles						
Serves as grade-level or subject area chair	15.9	20.3	-4.4	33.0	40.2	-7.2
Serves on school improvement committee	25.8	39.7	-13.8	43.8	43.0	0.7
Works to obtain external funding for the school	18.4	17.9	0.5	12.5	12.6	-0.1
Leads or promotes teacher collaboration	54.8	50.2	4.6	47.3	51.2	-3.8
Observes or provides feedback to other teachers	29.3	33.0	-3.7	43.2	48.8	-5.6
Involved in other activities	33.6	32.0	1.6	39.6	27.6	12.0
Sample Size[b]	**65**	**67**		**112**	**87**	

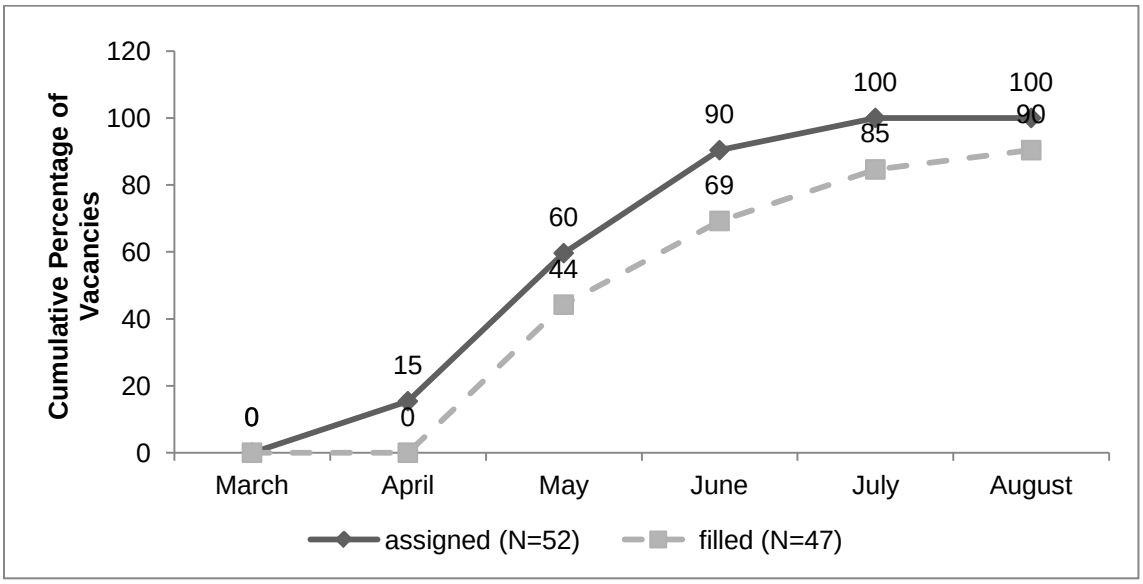
Source: 2009–2010 Mathematica Teacher Background Survey.

[a] Units are percentages unless otherwise indicated.

[b] Sample size is number of teachers.

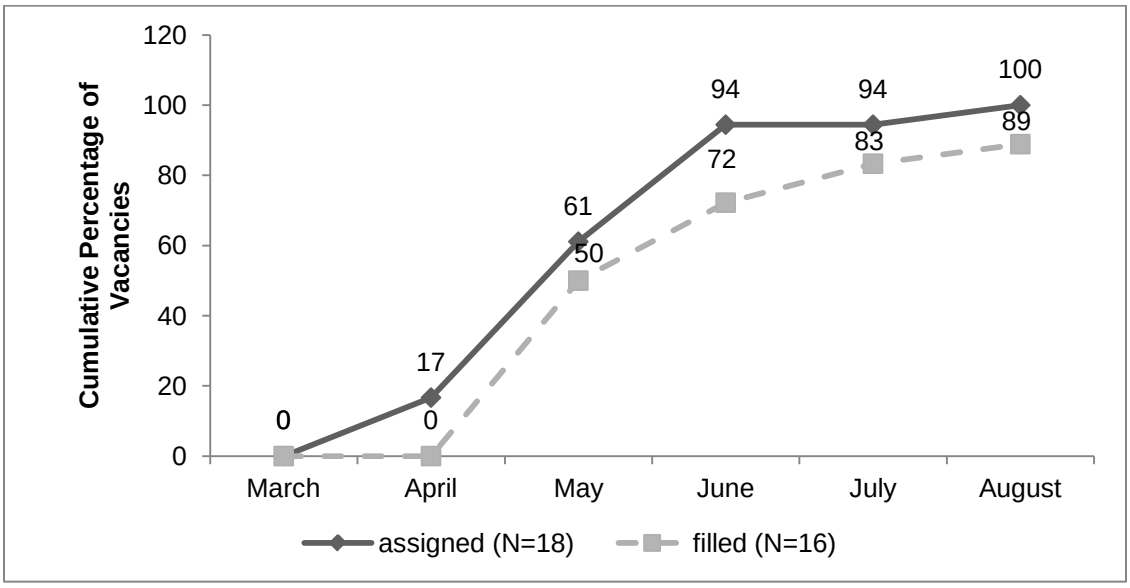
* Differences are statistically significant at the 0.05 level using a two-sided test. See Table IV.7 for results based on selective definition of focal teachers.

Figure C.1. Percentage of Elementary-Level TTI Vacancies Assigned and Filled by Month



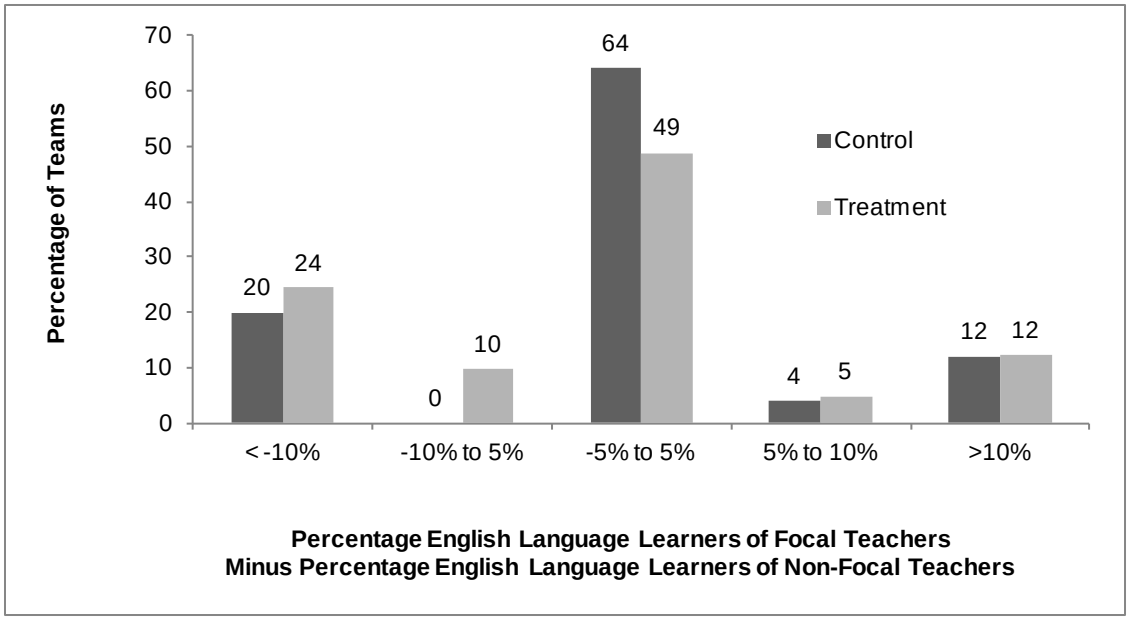
Source: TTI program records.

Figure C.2. Percentage of Middle School-Level TTI Vacancies Assigned and Filled by Month



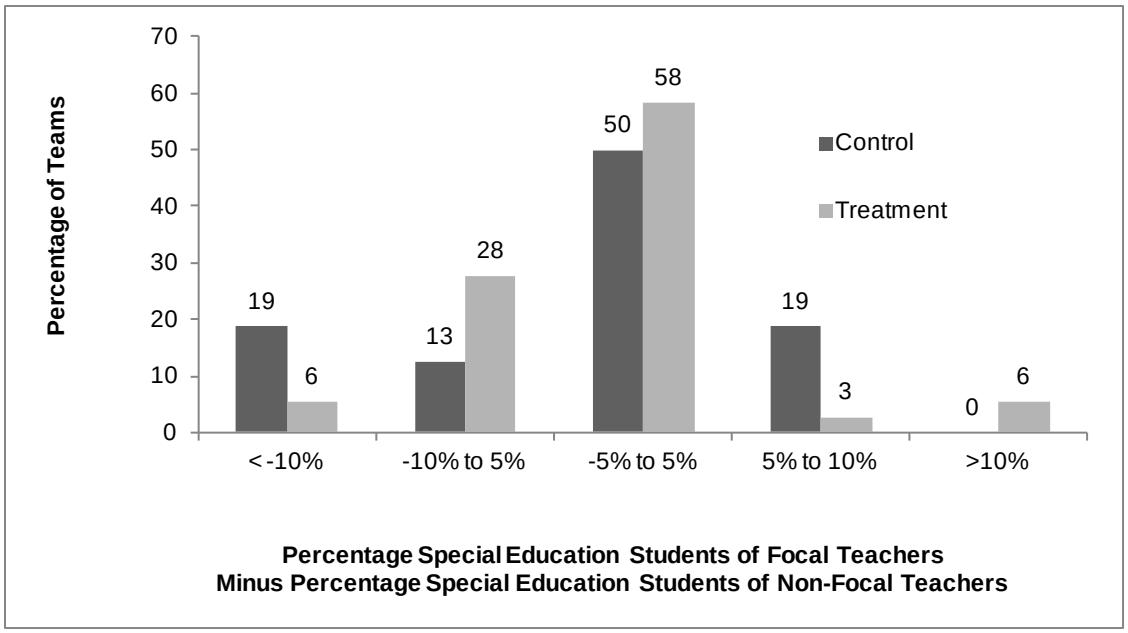
Source: TTI program records.

Figure C.3. Are Focal Teachers Assigned More English Language Learners Than Are Their Peers? Results by Treatment Status



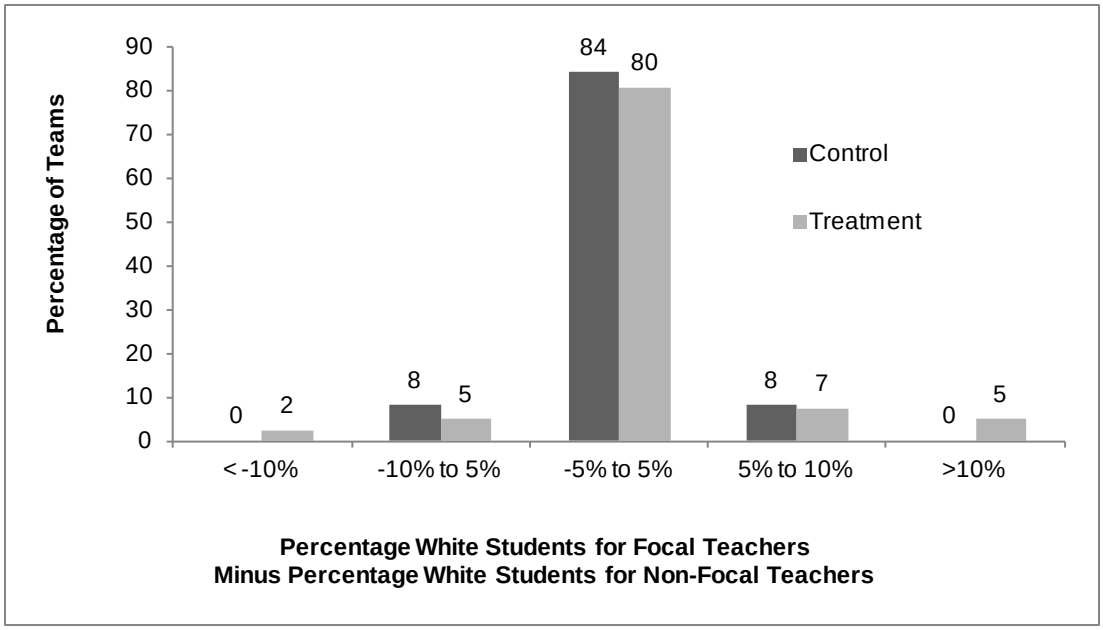
Source: Administrative data.

Figure C.4. Are Focal Teachers Assigned More Special Education Students Than Are Their Peers? Results by Treatment Status



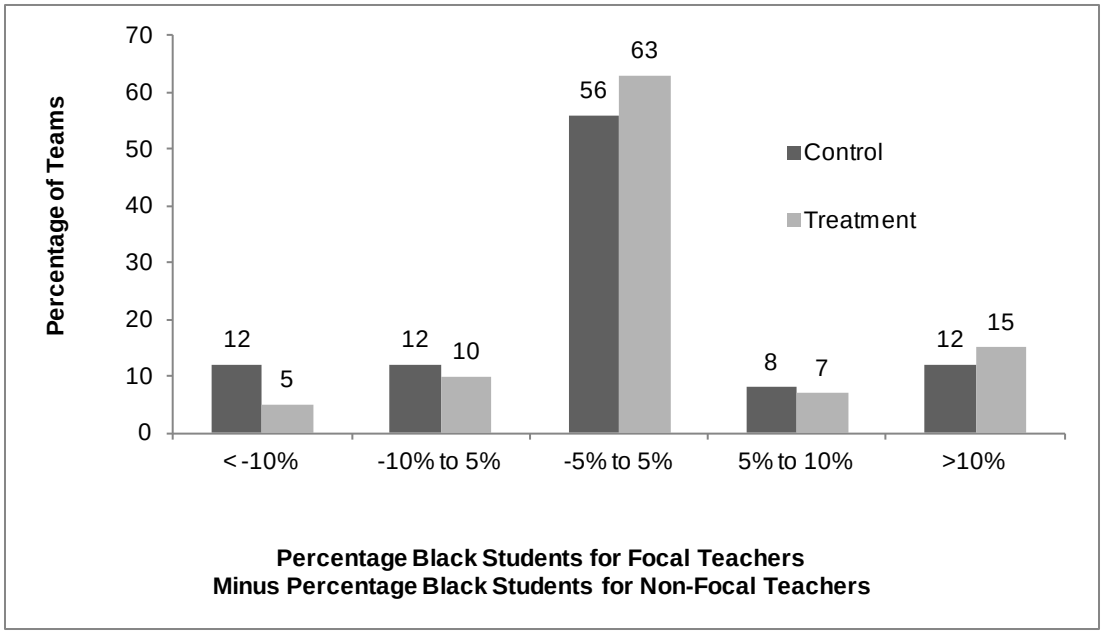
Source: Administrative data.

Figure C.5. Are Focal Teachers Assigned More White Students Than Are Their Peers? Results by Treatment Status



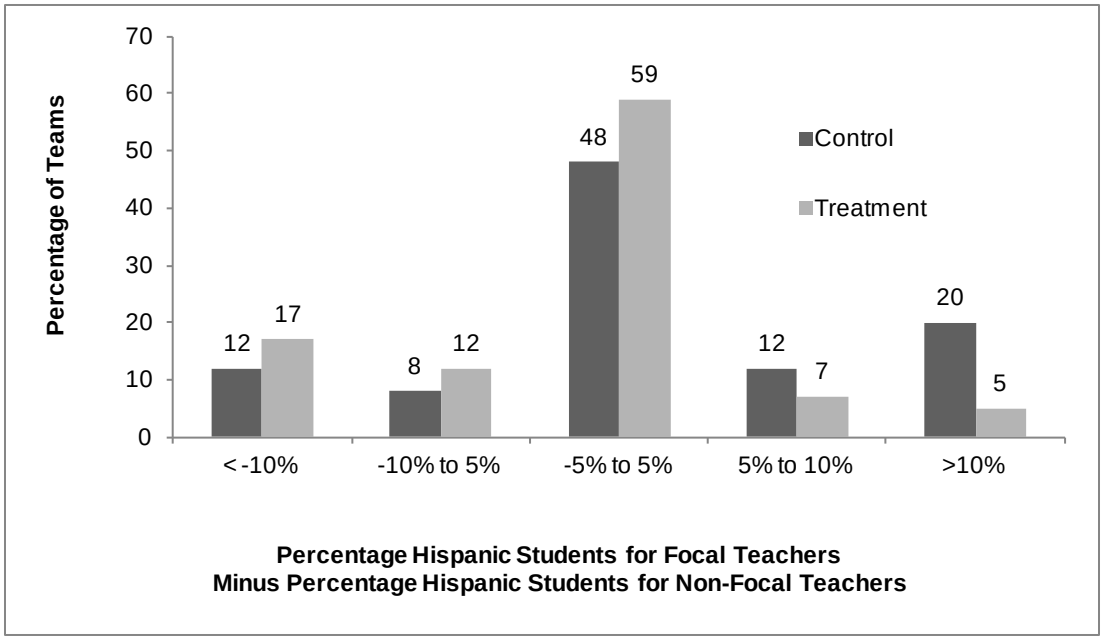
Source: Administrative data.

Figure C.6. Are Focal Teachers Assigned More Black Students Than Their Peers? Results by Treatment Status



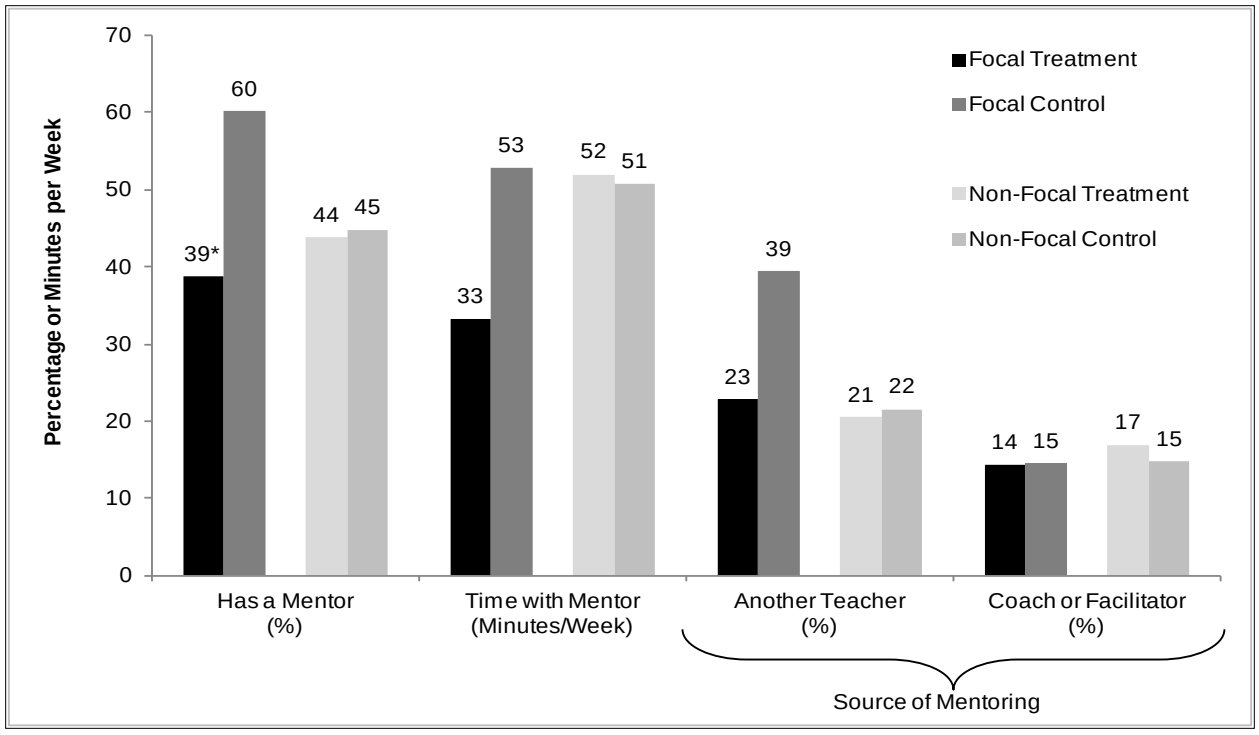
Source: Administrative data.

Figure C.7. Are Focal Teachers Assigned More Hispanic Students Than Are Their Peers? Results by Treatment Status



Source: Administrative data.

Figure C.8. Mentoring Received by Teachers, by Treatment and Focal Teacher Status, Using Inclusive Definition of Focal Teachers



Source: 2009–2010 Mathematica Teacher Background Survey.

Note: N = 64 treatment focal, 67 control focal, 112 treatment non-focal, and 87 control non-focal teachers. See Figure IV.6 for results based on selective definition.

* Treatment-control difference (within focal status) is statistically significant at the 0.05 level using a two-sided test.