Research Report

JONATHAN SCHWEIG, ELAINE LIN WANG, SABRINA LEE, KATA MIHALY

Teach For Pakistan Evaluation

Quantitative and Qualitative Study Findings

For more information on this publication, visit **www.rand.org/t/RRA1870-2**.

Published by the RAND Corporation, Santa Monica, Calif.

About This Report

The authors of this report present findings from a one-year mixed-methods study of the effect of the Teach For Pakistan (TFP) teacher leadership development program on whole-child development, perceptions of teaching quality, and perceptions of the contributions of TFP Fellows to the school community. This report contributes to a growing body of research on Teach For All organizations across the globe and the extent to which such leadership development and teacher training initiatives can enhance teaching and support whole-child development. The intended audience for this report includes policymakers, practitioners, and academics seeking to understand the impact and implementation of the Teach For All approach in a South Asian context.

This report borrows text from another RAND report, titled *Teach For Nigeria Evaluation: Quantitative and Qualitative Study Findings*; the description of the contextual background, the research questions, the limitations, and some of the recommendations are similar to what is presented in Mihaly et al. (2024).

RAND Education and Labor

This study was undertaken by RAND Education and Labor, a division of RAND that conducts research on early childhood through postsecondary education programs, workforce development, and programs and policies affecting workers, entrepreneurship, and financial literacy and decisionmaking. This study was sponsored by Teach For All with a grant from Porticus. Its content does not necessarily reflect the views or policies of the sponsor or funder.

More information about RAND can be found at www.rand.org. Questions about this report should be directed to kmihaly@rand.org, and questions about RAND Education and Labor should be directed to educationandlabor@rand.org.

Acknowledgments

We thank Murugi Kagotho and Robbie Dean at Teach For All and Tooba Akhtar, Khadija Bakhtiar, and Rabiah Chaudhry at TFP for their guidance and support at all stages of this study. We are grateful to all of the TFP staff who participated in the sense-making sessions and provided feedback about the initial qualitative study findings. We also thank the Research Consultants team for their efforts in collecting the baseline and follow-up data. We thank Irfan Ahmad for his leadership and Asad Ali Tabassam and Riaz Ahabib for overseeing the data collection and organizing activities to make the data collection a success. We extend our thanks to the enumerators and supervisors who visited schools, the data entry clerks and supervisors who ensured that the data were correctly coded, and the facilitators, the notetaker, and the transcriber who prepared the qualitative data for analysis. This study would not have been possible without the support of Pakistan's Ministry of Federal

Education and Professional Training and the Federal Directorate of Education leadership. We are grateful for the cooperation we received from the schools, principals, and area education officers who facilitated data collection. We are very thankful to all who participated in the data collection—students, parents, TFP Fellows, non-TFP teachers, and principals. We are grateful to Anne Fitzpatrick and Jason Lee for providing thoughtful guidance to improve this report. We are also grateful to Zohan Tariq for conducting quality checks of the audio recordings and transcripts and for assisting with the translation of the student survey into Urdu. Finally, we thank Yesenia Lizardi for editing assistance and our publications and editing team, including Monette Velasco and Nicole Bennett. Any flaws that remain in the report are solely our responsibility.

Summary

Substantial research has shown that classroom teachers play a critical role in shaping the lives of their students both academically and in terms of their social and emotional development (Nye, Konstantopoulos, and Hedges, 2004; Kunter et al., 2013; Fricke et al., 2021). Given this finding, initiatives that are focused on improving teacher recruitment, training, and development pipelines could be critical to improving schools and the lives of young people. RAND researchers conducted a one-year quantitative and qualitative evaluation of the Teach For Pakistan (TFP) leadership development and teacher training program. The goal of this evaluation was to examine the effect of the TFP program on whole-child development, perceptions of teaching quality, and perceptions of the contributions of TFP Fellows to the school community. The intended audience for this report includes policymakers, practitioners, and academics seeking to understand the impact and implementation of the Teach For All approach in a South Asian context. This report is part of a larger effort led by RAND to examine the short-term impact of Teach For All partners in two developing countries. As a part of this larger effort, RAND recently completed a two-year quantitative and qualitative study on the Teach For Nigeria program (Mihaly et al., 2024).

TFP focuses on addressing students' comprehensive needs, including academic achievement and key social and emotional competencies. As of this writing, there was relatively little evidence on the extent to which the program enhances the practices of teachers and supports student development. Specifically, no prior published work has examined a Teach For All network partner in the South Asian context.

This report presents findings from the one-year quantitative and qualitative studies (which took place between October 2023 and May 2024). In the quantitative study, we examined the following research questions:

1. What is the effect of TFP Fellows on whole-child development, including student academic achievement and social and emotional learning (SEL)?
2. What is the effect of TFP Fellows on students' and teachers' perceptions of teaching quality and the quality of the learning environment?
3. What is the effect of TFP Fellows on students', teachers', and principals' perceptions of school climate?

In the qualitative study, we examined the following research questions:

4. What do various stakeholders (e.g., principals, other teachers, parents, students, and TFP Fellows themselves) perceive as the contributions of TFP Fellows on whole-child development, including student academic learning and SEL outcomes?
5. How do stakeholders characterize the teaching approach and classroom environment of the TFP Fellows?
6. What do stakeholders perceive as the contributions of TFP Fellows to the school community?

The quantitative study sample included 80 principals, 162 teachers, and approximately 4,899 students from 80 government or public schools in the Islamabad Capital Territory. One-half of these schools employed at least two TFP Fellows (the treatment group), and one-half were comparison schools with no TFP Fellows that we selected because of their student demographic composition, size, and geographic proximity to treatment schools. Sixteen TFP schools were also selected for qualitative data collection.

To investigate the impact of the TFP program, we endeavored to measure students holistically, with respect to both academic achievement and social and emotional development. We administered assessments to students in mathematics, science, and English. To examine the impact of TFP on SEL, teaching quality, the quality of the learning environment, and perceptions of school climate, we administered a student survey, a teacher survey, and a principal survey. The student survey consisted of four SEL scales (empathy, growth mindset, self-management, and self-efficacy), five teaching and classroom conditions scales (control, challenge, rigorous expectations, emotional safety, and care), and three school climate scales (liking for school, safety, and school connectedness). We identified these constructs in collaboration with TFP, understanding them as the intended outcomes of focus for its programming. The teacher survey consisted of seven teaching and classroom conditions scales (clarity, cognitive activation, classroom management, assessment use, self-efficacy, responsibility for learning, and collaboration) and three school climate scales (teacher-student relationships, school climate, and physical safety). The principal survey consisted of two school climate scales (teaching quality and the school environment). A local data firm, Research Consultants (RCons), conducted all data collection. All assessment and survey data collection occurred in October 2023 (at baseline) and again in May 2024 (at follow-up). We used a pretest-posttest quasi-experimental design with systematically matched groups to examine the impact of TFP on a variety of outcomes.

To gain insights into TFP Fellows' contributions to students' development, classroom and instructional characteristics, and interactions with the school community, we collected qualitative data from 16 TFP schools in the study. These data came from interviews with TFP Fellows and principals and focus groups with non-TFP teachers and a sample of parents of children taught by TFP Fellows. Furthermore, we conducted brief interviews with a sample of students taught by TFP Fellows. To analyze the data from these interviews and focus groups, we conducted multiple rounds of applied thematic analysis across these focal schools.

Key Findings

Our analysis revealed the following key findings:

- Academic achievement improved more for students of TFP Fellows than for students of non-TFP teachers. Students in treatment classrooms scored between 0.15 and 0.41 standard deviations higher than students in comparison classrooms on mathematics, English, and science assessments.
- There was no quantitative evidence that SEL improved more for students of TFP Fellows than for students of non-TFP teachers. Similarly, quantitative estimates did not point to meaningful effects of the TFP program on perceptions of teaching quality and the quality of the learning

environment. There were also no significant differences in school climate for schools with TFP Fellows compared with schools with no TFP Fellows.

- Qualitatively, interviewees perceived that students of TFP Fellows have developed academically, particularly in English, mathematics, and critical thinking. They also reported observing improved grit and confidence, collective responsibility, and agency.
- In the qualitative study, stakeholders converged on several notable characteristics of TFP Fellows' teaching approach and classroom environment: TFP Fellows teach for mastery and conceptual understanding, motivate students to learn, cultivate a caring classroom environment, and empower students to lead.
- In the qualitative study, stakeholders perceived that TFP Fellows made school-level contributions, particularly in the areas of fostering parental engagement, ending corporal punishment, and shifting toward a student-centered approach to classroom management.

Recommendations

We make the following recommendations:

- **Consistent with existing policies, TFP should continue to encourage its teachers to engage in key practices**, such as fostering students' sense of self (e.g., grit and confidence), teaching students to collaborate, empowering them to lead, and engaging parents in their children's education. These practices were identified by multiple stakeholders as notable for improving whole-child outcomes.
- **TFP should consider implementing targeted SEL interventions and curricula with explicit instructions on SEL to improve the skills identified in the program's theory of change.** Prior research shows that explicit instruction is the most effective way to improve SEL skills, although TFP might wish to pilot-test such an intervention to ensure that it is appropriate for the Pakistani context.
- **TFP should use multiple modes to measure student outcomes and obtain a comprehensive understanding of student progress.** Self-reporting on SEL measures using surveys has known limitations. TFP might consider alternative data collection, such as performance-based SEL measures or third-party independent observations.
- **TFP should consider providing additional supports (e.g., programming, contextualized coaching) to help TFP Fellows navigate school contexts in which principals or non-TFP teachers might disagree with their methods.** These trainings can be extended to principals and non-TFP teachers.

Contents

Figures and Tables

Figures

Tables

Chapter 1

Introduction

Classroom teaching plays a critical role in shaping the lives of young people, both academically and in terms of their social and emotional development (Nye, Konstantopoulos, and Hedges, 2004; Kunter et al., 2013; Fricke et al., 2021). Given this finding, initiatives that are focused on improving teacher recruitment, training, and development pipelines could be critical to improving the lives of young people. One example of an alternative teacher preparation model that has been shown to be effective originated with Teach For America and has been adapted in more than 60 countries through the Teach For All network. The core program is a leadership development and teaching fellowship that recruits and prepares promising leaders who commit to teaching in lower-income schools for at least two years. The aspiration is that this training and work experience will develop a foundation among Teach For All network teachers to support a lifetime of working to positively affect marginalized children and communities. There is extensive, rigorous evidence that this type of teacher training program leads to improved student achievement (Clark et al., 2017; Lavado and Guzmán, 2020; McLean and Worth, 2023) and suggestive evidence that it improves student social and emotional learning (SEL) (Peña and Chacón, 2017). However, there has been no research conducted on the effect of the program in the context of developing countries in Asia.[1]

We conducted a one-year mixed-methods evaluation of the fellowship program of Teach For Pakistan (TFP), a Teach For All partner. The goal of the quantitative and qualitative evaluation was to examine the effect of the TFP program on whole-child development, perceptions of teaching quality, and perceptions of the contributions of TFP Fellows to the school community. The intended audience for this report includes policymakers, practitioners, and academics seeking to understand the impact and implementation of the Teach For All approach in a South Asian context. This evaluation is part of a larger effort led by RAND researchers to examine the impact of two Teach For All partners in developing countries. In early 2024, RAND completed a related quantitative and qualitative study on the Teach For Nigeria program (Mihaly et al., 2024).

Evaluation Objectives

The goal of our evaluation was to understand the one-year effects that TFP Fellows have on *whole-child development*, perceptions of *teaching quality*, and perceptions of the contributions of TFP Fellows to the *school community*. We investigated these issues through two studies that were conducted in parallel. Our quantitative study employed a quasi-experimental evaluation method; our qualitative study employed interviews and focus groups. These two studies were designed to provide different

[1] This chapter draws from material included in Chapter 1 of Mihaly et al. (2024).

perspectives on the impacts and contributions of TFP Fellows. In the quantitative study, we examined the following research questions:

1. What is the effect of TFP Fellows on whole-child development, including student academic achievement and SEL?
2. What is the effect of TFP Fellows on students' and teachers' perceptions of teaching quality and the quality of the learning environment?
3. What is the effect of TFP Fellows on students', teachers', and principals' perceptions of school climate?

In the qualitative study, we examined the following research questions:

4. What do various stakeholders (e.g., principals, other teachers, parents, students, and TFP Fellows themselves) perceive as the contributions of TFP Fellows on whole-child development, including student academic learning and SEL outcomes?
5. How do stakeholders characterize the teaching approach and classroom environment of the TFP Fellows?
6. What do stakeholders perceive as the contributions of TFP Fellows to the school community?

Evaluation Design Overview

This evaluation took place in four sectors of Islamabad Capital Territory, Pakistan, between October 2023 and May 2024. The quantitative and qualitative studies are intended to provide insights from multiple perspectives regarding the effect of TFP on whole-child development, as well as perceptions of teaching quality and perceptions of the contributions of TFP Fellows to the school community.

In preparation for the evaluation, we participated in a series of workshops facilitated by Teach For All with TFP leaders that were designed, in part, to help TFP articulate a program theory of change to guide the evaluation. We also hired Research Consultants (RCons), a Pakistan-based data-collection firm, to conduct school visits and collect all quantitative and qualitative data. We selected RCons as the data-collection firm for this activity through a competitive, rigorous request-for-quote process. Furthermore, we obtained Institutional Review Board approval for the data collection and approval to enter schools from the Federal Directorate of Education (FDE), the government agency that oversees public schools in the Islamabad Capital Territory. Prior to the data collection, RAND, TFP, Teach For All, and RCons conducted consultation meetings to review each of the evaluation instruments and related consent forms. We also developed and shared field manuals detailing procedures and guidance for both quantitative and qualitative data collection with RCons. RCons recruited and trained enumerators (i.e., data collectors) who visited schools to conduct the data collection from students, teachers, and principals.

For the quantitative study, we used a series of pretest-posttest quasi-experimental designs with systematically matched groups. Schools were identified as *treatment schools* if they had at least two TFP Fellows teaching in grades four, six, and seven. Then, each TFP school was matched to a *comparison school* (i.e., a school that had no TFP Fellows) similar in demographic composition, enrollment, grades served, and geographic location. A comparison teacher was selected to match a

TFP Fellow based on grade level and subject area taught (e.g., a TFP Fellow teaching grade-four English in a treatment school was matched to a grade-four English teacher in the comparison school). Then, one classroom per teacher participated in the student data collection. We applied covariate balancing methods, if necessary, to account for any remaining observable differences on key covariates that were associated with our outcomes. These covariate balancing methods were applied separately for each outcome, at each level of analysis, as appropriate. We matched 41 TFP schools with 39 comparison schools. The sample of TFP Fellows at baseline consisted of two Fellows in each school. Altogether, the TFP schools had 41 principals, 82 classrooms, and 2,694 students in grades four, six, and seven; the comparison schools had 39 principals, 80 classrooms, and 2,205 students in grades four, six, and seven.

Data collection for the quantitative study included four instruments: (1) grade-specific mathematics, English, and science assessments; (2) a student survey measuring social and emotional skills and competencies, student perceptions of teaching quality and school climate, and student background characteristics; (3) a teacher survey of Fellows and non-TFP teachers measuring school climate, teaching and learning, teacher well-being, and teacher background characteristics; and (4) a principal survey measuring school climate, the classroom environment of Fellows and non-TFP teachers targeted for the study, and principal background characteristics. Data collection was conducted in two waves: Baseline data were collected in October 2023,[2] and follow-up data were collected in May 2024.

The qualitative study included a total of 16 TFP schools, sampled purposively from the TFP schools in the quantitative study and in consultation with Teach For All and TFP. The organizations had particular interest in understanding the contributions of TFP Fellows for the full two years of their fellowships; in the second year, Fellows undertake the critical community partnership project (CPP). Given this, we decided that schools were eligible to participate in the qualitative study if they had at least two second-year Fellows.

Data collection for the qualitative study consisted of five protocols: (1) in-depth interviews with TFP Fellows, (2) in-depth interviews with principals, (3) focus groups with non-TFP teachers, (4) focus groups with parents of students taught by the TFP Fellows, and (5) brief interviews with students of TFP Fellows. We developed the protocols with input from Teach For All and TFP. We referenced the TFP program's theory of change and mapped interview and focus group questions onto the key evaluation questions. For example, protocol questions addressed perceived cognitive and noncognitive development in students, as well as TFP Fellows' classroom practices and environments, contributions to the school communities, and relationships with other educators, students, and parents. We collected interview and focus group data in May 2024. In all, we spoke with 15 principals, 36 TFP Fellows, 42 non-TFP teachers, 80 parents, and 64 students.

[2] We consider October 2023 to be the baseline for this study because it coincides with the study's launch. We cannot rule out the possibility that students in TFP schools had prior exposure to TFP Fellows. Because TFP Fellows work in a select set of schools, it is highly unlikely that comparison students or schools had prior exposure to TFP Fellows or the TFP program.

Limitations

We acknowledge various limitations to the data collection and methods employed in the quantitative and qualitative studies. First, we recognize that our sample might not be representative of all TFP schools. For the quantitative study, we collected data from 41 of the 62 schools in which the TFP program operated during the data collection period between October 2023 and May 2024. This sample might not be fully representative of all TFP schools in the country because it does not include schools in which TFP Fellows do not teach in grades four, six, or seven. Meanwhile, the qualitative study included only 16 schools, and we specifically sought to include schools that were in the quantitative study with at least two second-year Fellows. Findings from these schools might not generalize to all TFP schools. Moreover, though we gathered perspectives from different stakeholders in an effort to triangulate the qualitative data, our interviews included a limited number of non-TFP teachers and students.

Another limitation is that we relied on students' self-reports for most outcomes in our quantitative study, including student SEL, perceptions of teaching quality, perceptions of the classroom environment, and school climate. Researchers have reported various concerns with relying on student self-reports of these types of outcomes (Duckworth and Yeager, 2015; West et al., 2016), including reference bias (i.e., differences in self-reports might reflect differences in external frames of reference rather than true differences in skills), social desirability bias (i.e., students might inflate their responses to please others), and practice effects (i.e., scores on sequential surveys could be less accurate because of increased familiarity with the task). It is also possible that students misinterpreted the meaning of items—particularly in a setting, such as Pakistan, in which these constructs might not be familiar because many students have never taken a survey similar to the one that we conducted.[3]

In addition, we collected measures of school climate from only students and teachers of the classrooms selected for data collection; the resulting data might not represent the perceptions of the school as a whole. We also were unable to obtain balance between treatment and comparison groups on every measure of interest for the school-level analysis, even after using covariate balancing techniques. Therefore, these school-level results should be interpreted with caution.

Finally, though the quantitative and qualitative studies are complementary in that they provide different stakeholders' perspectives of student outcomes that can be triangulated, for multiple reasons, we expected there to be differences in findings from the quantitative and qualitative studies, and we are limited in the extent to which we can compare findings across the two sets of analyses. First, as previously noted, the sample of schools differed. We analyzed 41 treatment schools in the quantitative study. The quantitative study also consisted of information from 39 comparison schools in the analysis. Meanwhile, the qualitative study included only 16 treatment schools. Second, the sources of the SEL perceptions differed between the two studies. The quantitative study used student self-reports of social-emotional skills, whereas the qualitative study included parent, teacher, and principal reports in addition to student self-reports. Third, and importantly, the constructs in the data-collection instruments were not designed to be parallel. The student survey questions focused on specific aspects of a construct (e.g., self-efficacy) that could be measured with a few survey items and

[3] To mitigate these concerns, we conducted a pilot of the student data collection, and we revised items based on feedback from students.

were worded to help students understand these concepts. The interview and focus group data collection probed on some similar constructs, but it was much more open-ended, allowing respondents to comment on other aspects of whole-child development. Given this difference in the data-collection instruments, trying to reconcile a survey finding for school connectedness, for example, with the qualitative finding about teachers' relationships with students is ill-advised because school connectedness is a school-level construct, while teachers' relationships with students likely affect classroom climate and a sense of belonging at the classroom level. Along the same lines, we note that our qualitative investigation of TFP Fellows' contribution to the school community focuses on a different set of constructs from the quantitative measure of perceptions of school climate.

Organization of This Report

The remainder of this report is organized into four chapters. In Chapter 2, we describe the TFP fellowship program, including the program theory of change. Chapter 3 is devoted to the quantitative study. We describe our data sources, selection of schools, analytic methods, and results. In Chapter 4, we describe the qualitative data and our findings on the perceptions of various stakeholders about the contributions of TFP Fellows in 16 focal schools that were purposefully sampled from the 41 TFP schools in the quantitative study. Finally, in Chapter 5, we summarize our findings and make recommendations. We provide additional methodological details in the appendixes.

Chapter 2

The Teach For Pakistan Fellowship

Teach For All consists of a global network of more than 60 independent, locally governed partner organizations and a global organization that facilitates connection and learning to meet the mission of "ensur[ing] all children can fulfill their potential" (Teach For All, undated). Teach For All aims to enhance educational equity and outcomes by recruiting promising leaders to teach in underserved, high-needs schools for two years. During this time, Fellows work with stakeholders in the schools and broader communities to improve the quality of education and conditions necessary for students to prosper and become future leaders. The Teach For All program is intended to have both short- and long-term effects: In the short term, the goal is for teachers to positively influence their students, and, in the long term, alumni of the program are expected to continue working in education or related sectors.

There is a growing body of research on how to equip teachers to address student holistic development and learning more effectively (Grant et al., 2017; Cipriano et al., 2023), and these evidence-based practices guide Teach For All's Teaching as Collective Leadership framework. Teach For All's training is intended to support teachers' mindsets, skills, and knowledge to ensure students' holistic development and learning. Prior studies have found strong evidence that the program improves student academic learning (Clark et al., 2017; Mihaly et al., 2024) and suggestive evidence that it improves student social and emotional skills (Peña and Chacón, 2017; Kershner et al., 2019).[4] Additionally, a 2023 case study found that Teach For All's coaching and virtual workshops on how to use the framework were positively associated with the alignment of teacher mindsets with the Teaching as Collective Leadership framework (Teach For All, 2023).

In the remainder of this chapter, we describe the TFP fellowship program context and provide an overview of the program itself.

Teach For Pakistan Program Context

TFP joined the Teach For All network in 2018. The program grew out of a recognition of the scale of educational inequity in the country: About 20 million Pakistani children are out of school, about 40 percent of fifth graders cannot read in any language, and the vast majority of students attend struggling schools (TFP, undated-e). TFP sought to shape young talent into leaders who could help find solutions to the nation's education crisis. TFP's overarching vision is for the country's children to "grow into thinkers, scholars, artists, scientists, and entrepreneurs—builders of a new Pakistan" (TFP, undated-a).

[4] A full review of the literature supporting the Teach For All theory of change can be found in Chapter 2 of Mihaly et al. (2024).

Pakistan is a South Asian country, and its capital is Islamabad. Pakistan has the fifth largest population in the world (International Trade Administration, 2024) with more than 241 million people, about 95 percent of whom are Muslim. Urdu is the national language, although English is used often in official and administrative arenas and has been taught in schools since colonial times (Consulate General of Pakistan, undated). The ability to speak and use English is often associated with class, educational status, and expanded opportunities.

The United Nations International Children's Emergency Fund (UNICEF) reports that Pakistan "has the world's second-highest number of out-of-school children aged 5–6," which represents 44 percent of the population of this age group, and notes that gender disparities are large (UNICEF, Pakistan, undated). Pakistani girls, particularly in rural areas, are more affected by poor quality education than their male counterparts. They are less likely to enroll in school, to stay in school, and to attain learning outcomes, likely because of sociocultural beliefs that education is not essential for girls (World Bank, 2024).

The national education system consists of 12 years of education; grades one through five are referred to as *primary* or *elementary school*, and grades six through eight are referred to as *middle school*. At these levels, the curriculum comprises many subjects, including Urdu, English, mathematics, and science. Almost all public schools in Pakistan are single gender, whereas almost all private schools are mixed gender. About 35 percent of elementary students attend private schools. Many government schools lack essential infrastructure, such as electricity and drinking water. Teacher absenteeism and a lack of qualified and trained teachers are also barriers to quality public education (World Bank, 2024). One estimate suggests that 14 percent of public school teachers are absent on any given day and that many teachers lack sufficient content knowledge to pass examinations in the subjects they teach (World Bank, 2024).

Pakistani teachers receive limited preservice training (two to three days), and, in some cases, teachers are placed in schools without any training. Continuous professional development varies in each region, and a limited number of teachers typically receive this training for six to nine days per year. Training includes such topics as curriculum, lesson planning, and subject-specific pedagogy. Teachers generally do not prepare lesson plans, and systems for monitoring teachers are generally weak. Teachers are responsible for teaching a mix of grades and subjects, with up to eight subject and grade combinations (e.g., English and science taught to students in grades two, three, four, and five).

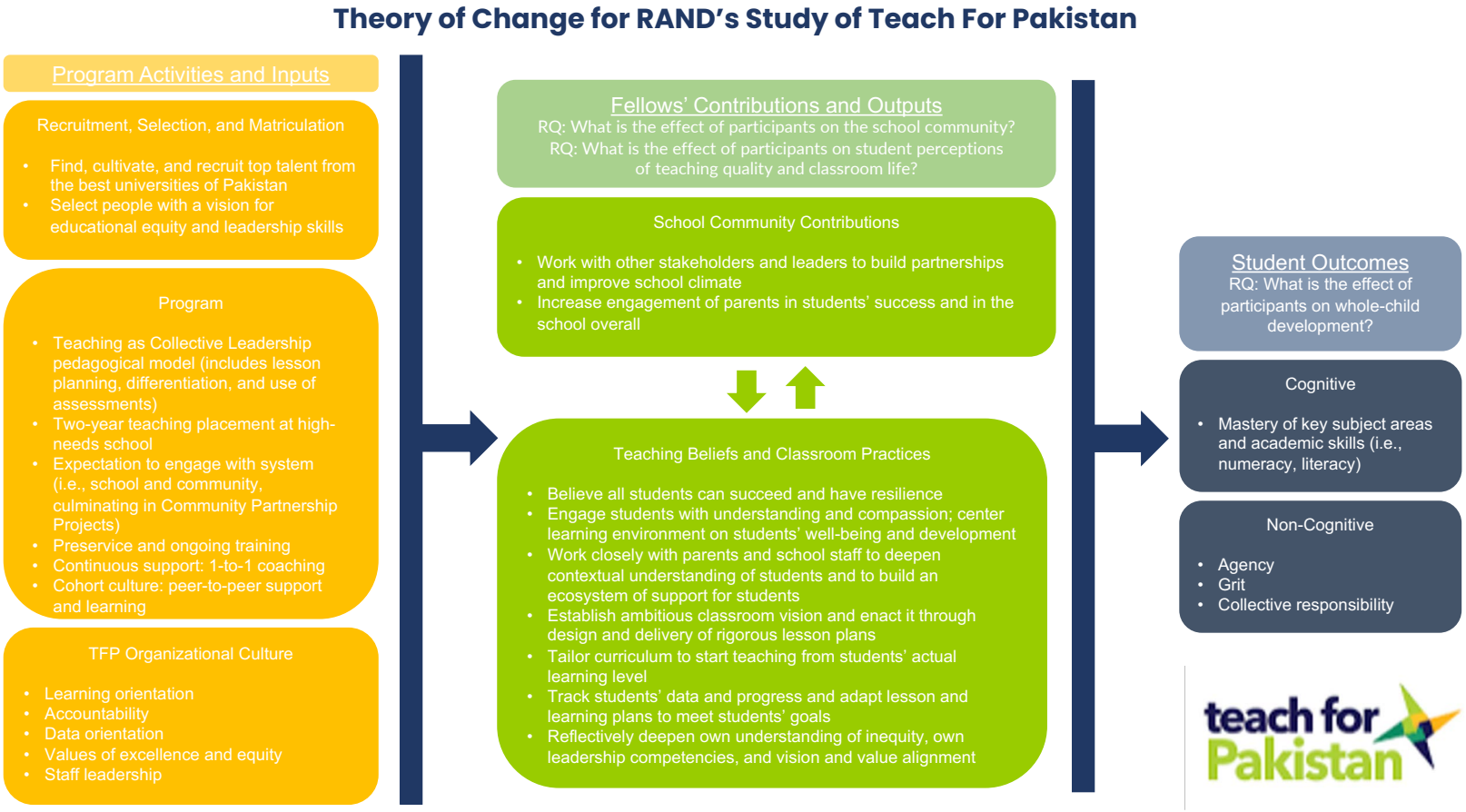
The Pakistani school year begins around March and continues through June, after which there is a summer break during monsoon season, and then school picks up again in August and runs through February. In several provinces, including the territory in this study, at the end of grades five and eight, students must take standardized examinations. The results are reported to the provinces' respective departments of education, and there are requirements for proceeding to the next level of education. Given these requirements, many students are held back in grades four and seven to allow them additional time to learn the skills and content necessary to pass these examinations. Students might be held back in other grades, or they might be promoted to the next grade without having mastered the grade-level content. TFP leaders report that it is rather common for students to be up to four levels behind their grade level (e.g., for a grade-four student to lack grade-one skills and content mastery). The pass rate on the standardized examinations and the teacher attendance rate are key indicators of success for Pakistani schools. Pressure is high on teachers to have their students do well. Teachers,

therefore, typically focus on grade-level exam content and have students memorize such content. Support for teachers to apply less traditional pedagogical strategies—for example, active, hands-on, or inquiry-based learning—is low.

Teach For Pakistan Program Overview

The TFP fellowship program is designed to recruit exceptional university graduates and young professionals from a variety of fields and develop them into "lifelong leaders and advocates of systemic change in the educational system" (TFP, undated-d). Recruitment and selection is the first key program activity or input reflected in TFP's theory of change, which was developed for this study (see Figure 2.1). Historically, TFP Fellows have come from diverse backgrounds (e.g., engineering, finance, economics, social sciences) and often have not had any teaching or education experience. In the spring of 2022, prior to the start of the evaluation period, there were more than 3,300 applicants to the program; only 3 percent were accepted and invited to join the cohort (TFP, 2024). The Fellows in that cohort were from 35 cities and 20 universities, about half of the cohort were from the top seven universities in Pakistan, and almost two-thirds were female (TFP, 2024). In 2024, the program was working with 165 Fellows, placed in 62 government schools. Since the program's establishment in 2018, 302 Fellows have completed their two-year teaching commitments (TFP, undated-b).

Figure 2.1. Teach For Pakistan's Theory of Change



SOURCE: TFP, 2023.

The two-year, full-time, paid fellowship begins by training the recruits in teaching, leadership, and community mobilization (TFP, undated-d). They receive grounding in key principles of leadership aligned with TFP's theory of change. Specifically, the four focal areas of Fellows' development are self-leadership, classroom leadership, collective leadership (i.e., the ability to mobilize people for community and collective action), and system leadership.

TFP prepares Fellows for their teaching placement in a six-week intensive training institute. Fellows engage with a variety of pedagogical topics, including planning lessons, differentiating instruction for students with diverse needs, using assessments to inform instructional next steps, and managing classrooms (Figure 2.1). The training also helps Fellows set and own a vision of outcomes for students in line with the overarching priorities of TFP and then supports them to acquire and apply strategies to achieve that vision; in this sense, TFP Fellows have agency and exercise leadership. Driven by their visions, Fellows draft goals—for example, goals directed toward student comportment and classroom practices and culture—that they then work to fulfill. They also learn about the importance of co-creating goals with students and parents to ensure that both are invested in the goals. The training exposes Fellows to best practices culled from TFP and Teach For All experiences that they can use to intentionally plan their teaching approaches and interactions. For example, Fellows might learn that teaching students for mastery of content and skills involves frequent collection and examination of data to identify where students might need additional support and that more learning opportunities can be added before or after the school day.

Training also addresses students' nonacademic development. TFP imparts a vision of the classroom as a student-led space in which students take ownership of their and each other's learning and the teacher serves as a guide. TFP aims for its Fellows to enact practices that center love and support in helping students learn and grow. According to TFP program leaders, the program aspires to help students "invest in visions of their own selves" and "build a sense of possibility in students."[5] This includes helping students perceive and challenge injustice and systems of inequity. As one TFP leader said, "We want kids to engage in the conversation of, when they're behind [academically], it's because of a system that has kept them back, not due to their individual weakness." The program encourages Fellows to build classroom environments that enable students to develop these outlooks and dispositions and thrive. It shares and models explicit strategies for building students' sense of agency, grit, and collective responsibility.

After the initial training, the Fellows are placed in schools that serve children in low-income and underserved communities. TFP Fellows typically teach three primary subjects (English, mathematics, and science). Moreover, TFP's negotiated contract with schools allows Fellows to enact academic plans that might differ from the prescribed pacing that non-TFP teachers are required to follow, although Fellows are also held to account for students' performance on standardized examinations. TFP Fellows might, therefore, provide remediation, going back to ensure that students master foundational skills that they missed in past years.

For approximately the first three months of their field experience, TFP Fellows typically focus on diagnosing and understanding students' academic needs and developing academic plans to help students approach mastery of grade-level skills and content. Then, they spend the remaining time

[5] TFP program leadership, virtual communication with the authors, October 10, 2024.

implementing their plans. Similarly, Fellows take time to get to know their students in a holistic way (apart from academics), such as where they come from and their social and emotional needs. Fellows typically use such information to build their classroom culture and values to support students' social, emotional, and leadership development. They also distribute a letter to parents from TFP that states their stance against corporal punishment, a typical practice in Pakistani schools.

TFP Fellows receive a variety of continued support during their field experience. TFP coaches conduct at least four cycles of one-on-one classroom observations and debriefs, offering reflections and feedback on classroom management and lesson planning and delivery, among other topics. Coaches also facilitate five group meets, or communities of practice, with the Fellows in their charge. If they identify a particular need, coaches will suggest extra workshops to support a small group of Fellows to tackle emergent issues. TFP's leadership and training team supports Fellows by conducting at least seven cohort-wide training sessions and conducting small-group workshops as needed. It also shares learning resources (e.g., research papers, videos). In addition, Fellows participate in at least five informal peer-to-peer support gatherings designed to strengthen the relationships among Fellows and cohorts (TFP, undated-d).

TFP Fellows are expected to contribute beyond the classroom by developing relationships with the school community and spearheading a CPP in their second year of placement. TFP positions its Fellows as change agents whose impact ought to outlast the two years of their placement at the school. Partnering with parents and creating a culture of collaboration among teachers, for example, help awaken parts of the educational environment to better support students long term. The second-year CPP is an opportunity for Fellows to expand their leadership to address a system-level challenge they have identified as a barrier to student learning. With the support of TFP programming, Fellows design, fundraise for, and execute the project, activating stakeholders as needed. Past projects have addressed gaps in career counseling, digital and financial literacy, mental health, and reproductive health awareness for students (TFP, 2024).

Post-fellowship, TFP Fellows become TFP alumni with access to opportunities for continued professional growth and development that support them to be "change-makers" (TFP, 2024). Fellows do not automatically receive teaching positions after their fellowships. If they choose to pursue teaching careers, they must apply to the Ministry of Education, as all other prospective teachers do. A vast majority of TFP Fellows go on to work in and around education; they remain committed to transforming the education system in Pakistan.

According to TFP, Fellows "develop students' emotional and intellectual capacity, as well as their ability to act on their learning, solving real-life challenges" (TFP, undated-c). Specifically, TFP expects students of Fellows to grow in mastery in key subject areas and academic skills and to demonstrate the key noncognitive traits of agency (autonomy and ownership of one's goals and actions), grit (perseverance), and collective responsibility (regard for the well-being and development of others in the community).

Chapter 3

Quantitative Study of Teach For Pakistan

As described in Chapter 1, the quantitative study was designed to address the following research questions, which explore the impacts that TFP Fellows have on *whole-child development*, students' perceptions of *teaching quality*, and students' and teachers' perceptions of the *school climate*:

- What is the effect of TFP Fellows on whole-child development, including student academic achievement and SEL?
- What is the effect of TFP Fellows on students' and teachers' perceptions of teaching quality and the quality of the learning environment?
- What is the effect of TFP Fellows on students', teachers', and principals' perceptions of school climate?

Study designs that employ randomization are commonly regarded as the gold standard in education research because they have strong internal validity and support rigorous causal inferences (Burtless, 1995). The randomization process mitigates selection bias by ensuring that the treatment and comparison units (e.g., students, teachers, schools) are similar at baseline among all observed and unobserved variables that could influence the study outcomes. However, randomization was not feasible for this study. In terms of the students, randomization is impracticable because of the administrative complexity involved with securing permissions to randomly assign students to classrooms, which would require, among other things, agreements from local government. In terms of the teachers, randomization was not possible because it would have disrupted the way TFP engages with schools and purposefully places its Fellows in classrooms with the highest need.[6]

Instead, we used a pretest-posttest quasi-experimental design with systematically matched groups to estimate TFP impacts, and we applied covariate balancing methods to account for any remaining observable differences on key covariates associated with our outcomes to ensure that the treatment and comparison units were similar at baseline among observed variables. Although the threat of selection bias cannot be mitigated entirely, this research approach provides the most-rigorous causal evidence possible given the context of our study. For *student-level* outcomes (which include academic achievement and SEL as a part of research question 1), we consider students as *treated* if they were assigned to a TFP teacher in October 2023. For *teacher-level* outcomes (which include teaching quality and the quality of the learning environment in research question 2), we consider teachers as *treated* if

[6] This chapter draws from material included in Chapter 3 of Mihaly et al. (2024).

they were TFP Fellows in October 2023. For *school-level* outcomes (including school climate in research question 3), we consider a school as *treated* if it employed a TFP Fellow in October 2023.[7]

In this chapter, we describe the data sources, school selection, sample, analytic methods, and results from the quantitative study.

Data Sources

Data for the quantitative study came from surveys of students, teachers, and school principals, as well as assessments of students collected by the data-collection firm RCons. All surveys were pilot-tested in schools that were not included in the study to ensure that the wording of items was clear. Data collectors were trained on the instruments and the procedures for visiting schools, and approvals were obtained from the local government for the school visits. Data entry clerks were trained to ensure that identifiers were unique to respondents and to allow the two rounds of data collected to be linked. During data collection, RAND received daily reports of the number of schools visited and any concerns identified.[8]

Student Assessments

The mathematics, English, and science assessments were designed by TFP to assess student learning. These assessments were based on content standards and student learning objectives articulated in Pakistan's National Curriculum (Ministry of Federal Education and Professional Training, 2022). All assessments were fielded in English (which is the language of instruction) in paper format. The mathematics assessments included such topics as numbers and operations, measurement and geometry, and algebraic thinking. The English assessments included reading comprehension, grammar, phonics, and verbal reasoning. The science assessments included topics from Earth science, physical science, and the life sciences. The baseline assessments measured concepts from the prior school year, and the assessments administered at the first follow-up were aligned with the curriculum for the current grade of the student.[9] Mathematics, English, and science scores were generated for each student at baseline and follow-up by dividing the number of points earned by the total number of available points for an assessment, treating missing items as incorrect.

Student Surveys

Student surveys were developed to measure features that are aligned with TFP's theory of change: student social and emotional competencies (empathy, growth mindset, self-management,

[7] For student-level analyses, our design effectively simulates a randomized control trial (RCT) with students randomly assigned to TFP Fellows. For teacher-level analyses, our design simulates an RCT with teachers randomly assigned to the TFP fellowship program. For school-level analyses, our design simulates an RCT with schools randomly assigned to partner with TFP.

[8] The data-collection procedures in Pakistan were similar to those employed in Nigeria. Details of these procedures are described in Appendix B of Mihaly et al. (2024).

[9] For example, a student in grade four was given the grade-three mathematics assessment at baseline and the grade-four mathematics assessment at follow-up.

and self-efficacy), aspects of teaching and classroom conditions (control, challenge, rigorous expectations, emotional safety, and care), and aspects of school climate (liking for school, safety, and school connectedness). The surveys were also used to collect information about student background characteristics (gender, age, grade, home language, home possessions, parental education, household size, living arrangement, and time it takes to travel to school). The survey was adapted from several instruments, including the Panorama Student Survey (Panorama Education, 2015), the California Office to Reform Education (CORE) Districts SEL Survey (West et al., 2018), the 5Essentials Survey (Chicago Consortium on School Research, 2011), the Pakistan National Nutrition Survey (UNICEF, Pakistan, 2018), the Adolescent Measure of Empathy and Sympathy (Vossen, Piotrowski, and Valkenburg, 2015), the New General Self-Efficacy Scale (Chen, Gully, and Eden, 2001), the Tripod Survey (Ferguson, 2010), the Student Engagement in Schools Questionnaire (Hart, Stewart, and Jimerson, 2011), the California Healthy Kids Survey (Austin and Duerr, 2005), and the U.S. Department of Education School Climate Survey (National Center for Education Statistics, undated-a). To minimize survey development time and mitigate the possibility of over-alignment between the TFP program and the desired outcomes, intact scales were taken from each source as often as possible. Surveys were piloted with students, teachers, and principals, and, using feedback from the pilots, wording changes were made to several items so that they were more appropriate for the local Pakistani context.

Scale scores were created by averaging items related to measured social and emotional skills and competencies and students' perceptions of both teaching quality and school quality. Table 3.1 details reliability estimates (which we assessed by using Cronbach's alpha)[10] from the baseline survey administration. Most of the internal consistency estimates for the student-level constructs are above 0.70 and show good reliability.[11] The safety scale had the lowest estimated reliability, which was well below the threshold for acceptable reliability adopted in this study.[12] However, because this construct was considered important to TFP's theory of change, we included it in our analysis.

[10] Cronbach's alpha is a widely reported measure of a kind of reliability known as *internal consistency*. It compares shared variance among items in a scale with the total observed variance.

[11] We adopt conventions used by Organisation for Economic Co-operation and Development (OECD) (OECD, 2008; OECD, 2013). We interpret a Cronbach's alpha value greater than 0.70 as evidence of good reliability and a value greater than 0.50 as adequate reliability.

[12] Low reliability can compromise the validity of inferences, including inferences about the association of safety with TFP (the treatment).

Table 3.1. Student Survey Constructs

Construct	Source	Illustrative Item	Items	Alpha
Empathy	Adolescent Measure of Empathy and Sympathy	I can easily tell how others are feeling.	8	0.68
Growth mindset	CORE Districts SEL Survey	I can do well in a subject even if I am not naturally good at it.	4	0.71
Self-management	CORE Districts SEL Survey	During the past 30 days, I got my work done right away instead of waiting until the last minute (i.e., finished my work before the deadline of the teacher).	4	0.68
Self-efficacy	New General Self-Efficacy Scale	I believe I can succeed at most any endeavor to which I set my mind.	8	0.85
Control	Tripod Survey	Student behavior in this class is a problem.	7	0.55
Challenge	Tripod Survey	This class requires me to work hard to do well.	4	0.63
Rigorous expectations	Panorama Student Survey	How much does your teacher encourage you to do your best?	5	0.78
Emotional safety	U.S. Department of Education School Climate Survey	In this class, students work on listening to others to understand what they are trying to say.	4	0.75
Care	Tripod Survey	My teacher in this class makes me feel that s/he really cares about me.	3	0.66
Liking for school	Student Engagement in Schools Questionnaire	Most mornings, I look forward to going to school.	4	0.81
Safety	5Essentials	I feel safe at this school.	4	0.50
School connections	California Healthy Kids Survey	There is a teacher in my school who checks on how I am feeling.	7	0.82

To minimize response burdens, the survey was split into two different forms (form A and form B), and one-half of the students were randomly assigned to complete each form. Both forms included the student background questions, and both included scales from the three research areas.[13] The survey was fielded in Urdu.[14] To ensure improved response rates, the student survey was administered in a classroom setting with an enumerator (i.e., a data collector) who read the survey items and response options out loud for students in grade four (students in grades six and seven completed the survey on their own). After hearing the items and responses, students who were read the questions recorded their answers on paper forms.

Teacher Surveys

The teacher surveys were developed to measure features that are aligned with TFP's theory of change: teaching quality (clarity, cognitive activation, classroom management, and assessment use), teacher self-efficacy (collaboration and responsibility for learning), school context (teacher-student relationships, school climate, and physical safety), and teacher social and emotional well-being. The surveys were also used to collect individual background characteristics (e.g., gender, formal teaching qualifications, teaching experience). The survey was adapted from several instruments, including the 5Essentials Survey (Bryk et al., 2010), the Schools and Staffing Survey (National Center for Education Statistics, undated-b), the Teaching and Learning International Survey (TALIS; OECD, 2018), the Young Lives Survey (Moore, 2016), the Alaska School Climate Survey (Spier, 2016), the Learning in the 21st Century Survey (Digital Promise, 2021), the Child Friendly Schools Survey (UNICEF, 2009), and the Annual Status of Education Report Pakistan (ASER Pakistan, undated). The surveys were administered by an enumerator who recorded responses on a tablet.

Scale scores were created by averaging items. Table 3.2 details internal consistency estimates from the baseline survey administration. Most of the internal consistency estimates are above 0.70, which shows adequate or good reliability (OECD, 2013). The assessment use scale had the lowest estimated reliability.

[13] Form A consisted of the following constructs: *growth mindset* and *self-efficacy* to measure student social and emotional competencies; *challenge*, *rigorous expectations*, and *emotional safety* to measure teaching and classroom conditions; and *connection* to measure school climate. Form B consisted of the following constructs: *empathy* and *self-management* to measure student social and emotional competencies, *control* and *care* to measure teaching and classroom conditions, and *safety* and *liking for school* to measure school climate.

[14] Although English is the language of instruction, the survey was administered in Urdu because our local data collectors believed that, based on their past experiences administering surveys to school-aged children in Pakistan, the concepts of the survey would be better understood in Urdu.

Table 3.2. Teacher Survey Constructs

Construct	Source	Illustrative Item	Items	Alpha
Well-being	TALIS 2018	In your experience as a teacher at this school, to what extent does your job negatively impact your mental health?	4	0.76
Clarity	TALIS 2018	Thinking about your teaching overall, how often do you present a summary of recently learned content?	4	0.71
Cognitive activation	TALIS 2018	Thinking about your teaching overall, how often do you give tasks that require students to think critically?	3	0.64
Classroom management	TALIS 2018	Thinking about your teaching overall, how often do you tell students to follow classroom rules?	5	0.69
Assessment use	TALIS 2018	How frequently do you review test or exam data with teachers in your grade level?	4	0.60
Self-efficacy	TALIS 2018	In your teaching, to what extent can you craft good questions for students?	11	0.76
Responsibility for learning	Young Lives	If I really try hard, I can get through to even the most difficult or unmotivated students.	19	0.76
Collaboration	TALIS 2018	On average, how often do you exchange or develop teaching materials with colleagues?	8	0.76
Teacher-student relationships	TALIS 2018	Teachers and students usually get on well with each other.	4	0.71
School climate	Alaska School Climate Survey	To what extent is each of the following a problem in *this* school? [An example problem:] Students deliberately missing classes.	14	0.87
Physical safety	5Essentials	My students are safe . . . in my classroom.	4	0.69

Principal Survey

The principal survey measured school climate (school environment and teaching quality) and collected individual background characteristics (e.g., gender, school leadership experience). The survey was adapted from several instruments, including the 5Essentials Survey (Bryk et al., 2010), the Teach For All Principal Survey, the Schools and Staffing Survey (National Center for Education Statistics, undated-b), TALIS 2018 (OECD, 2018), the Panorama Teacher Survey (Panorama Education, 2015), the Learning in the 21st Century Survey (Digital Promise, 2021), the Child Friendly Schools

Survey (UNICEF, 2009), and the Annual Status of Education Report Pakistan (ASER Pakistan, undated). The surveys were administered by an enumerator who recorded responses on a tablet.

Intact scales were taken from each source as often as possible. Scale scores were created by averaging items. Table 3.3 details internal consistency estimates from the baseline survey administration.

Table 3.3. Principal Survey Constructs

Construct	Source	Illustrative Item	Items	Alpha
School environment	Teach For All Principal Survey	The [TFP Fellow or the target teacher] promotes a positive, collaborative professional culture in the school.	9	0.91
Teaching quality	Teach For All Principal Survey	When thinking about [TFP Fellows or the target teachers], to what extent do they create classroom environments that engage students in challenging problems, texts, and/or questions?	12	0.87

School Selection

Selection of Teach For Pakistan Schools

The student populations of interest for the study are students who attend schools with TFP Fellows and students who attend comparison schools in which there are no TFP Fellows.[15] Through consultation with TFP and Teach For All, it was determined that teachers have the largest impact on whole-child development when students are in primary school. However, because completing surveys requires basic literacy and comprehension skills, the sample was restricted to students in upper primary school. For these reasons, and following discussions with Teach For All and TFP, we decided to restrict data collection to schools serving grades four, six, and seven in four sectors of the Islamabad Capital Territory: Bhara Kahu, Nirole, Sihala, and Tarnol. The grade ranges were selected based on a high concentration of TFP Fellow placements in these grades, partially because of the importance of country-wide assessments in grades five and eight.

TFP Fellows work in 62 schools in the Islamabad Capital Territory, and all of these schools are in the four study sectors. Treatment schools selected for the study are a subsample of these schools. Using power calculations and resource constraints related to the number of students, teachers, and principals that we could survey in the limited data-collection time frame, we determined that we needed 80 schools to participate in the study. Treatment schools were purposefully selected by the TFP team based on the grade level taught (four, six, or seven) and the number of TFP Fellows

[15] We refer to schools with TFP Fellows as *treatment schools*. We refer to schools with no TFP Fellows that were matched to TFP schools as *comparison schools*, and we refer to teachers in these school as *comparison teachers*.

teaching in each school (at least two teachers per school teaching the relevant grade levels).[16] In a few instances, three TFP Fellows taught the selected grade levels.

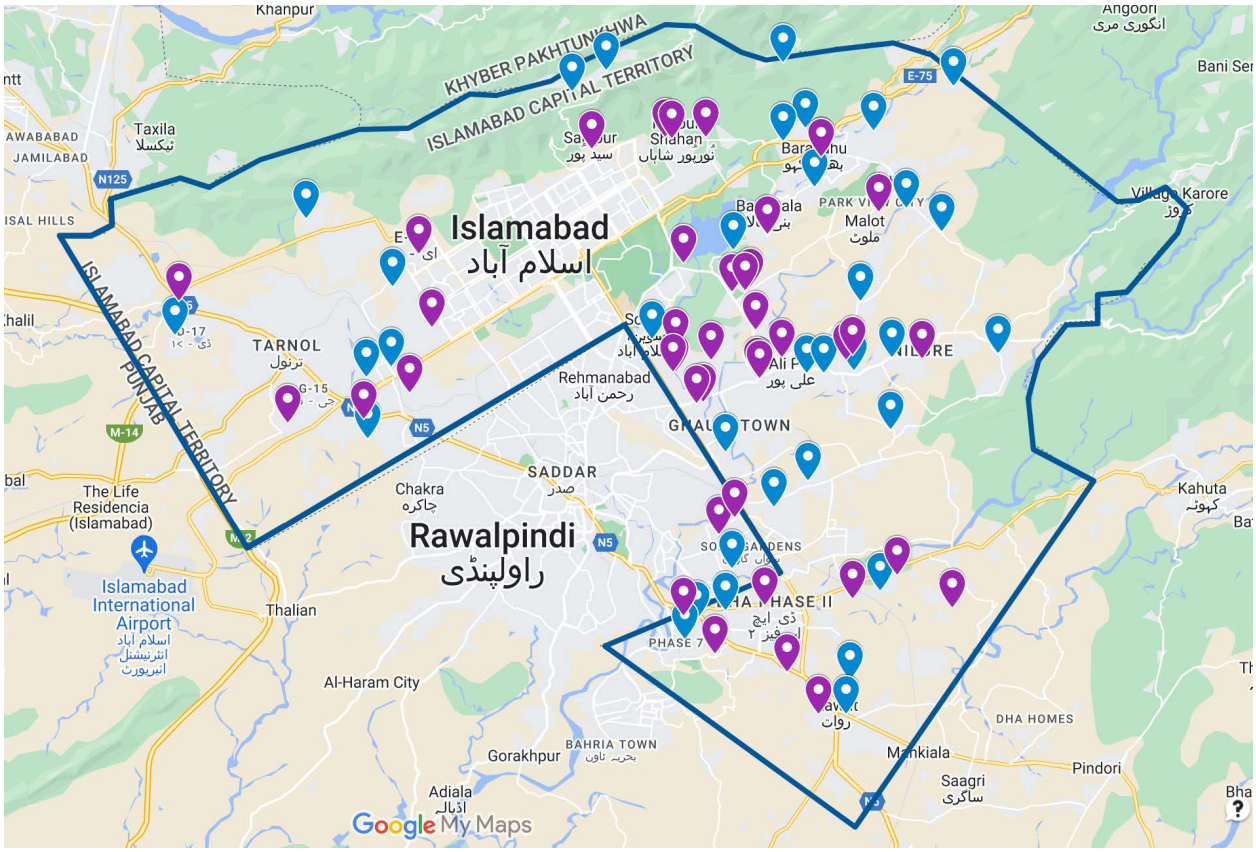
Selection of Comparison Schools

We took multiple steps to select the comparison schools. First, we obtained a list of all schools in the four sectors from FDE, the government agency that oversees public schools in the Islamabad Capital Territory. The information provided by FDE included the address, the grades served by the school, enrollment data, and the gender of the students (female, male, or coed). Early in the school year, TFP identified 42 potential treatment schools based on the grades and subjects taught by the TFP Fellows in the schools.[17] Using the FDE information, we selected 62 potential comparison schools (45 likely comparison schools and 17 backup schools). RCons then visited all 104 schools to confirm the data in the FDE file and collect information about the principals, teachers, and grades taught. In addition, teachers in comparison schools were identified for the study. Target teachers in comparison schools were selected because they taught the same grade levels as teachers in the matched TFP schools. After the initial visit by RCons, we revised the comparison school list, matching one comparison school to each treatment school. In the end, we identified a total of 82 schools for the study. Two comparison schools did not participate in follow-up data collection, leaving 41 treatment schools and 39 comparison schools in the study. The specific locations of the treatment and comparison schools are shown in Figure 3.1.

[16] There were a handful of schools with more than two TFP Fellows. In these cases, we purposefully selected teachers who taught grades or subjects that were less represented in the rest of the sample. For example, if a school had two sixth-grade TFP Fellows and two fourth-grade TFP Fellows, then we selected the two sixth-grade TFP Fellows for the study because there were fewer teachers in the rest of the sample teaching grade six.

[17] Of the 62 TFP schools, there were only 45 schools that had at least two teachers who taught English, mathematics, or science in grades four, six, or seven. TFP selected 42 of these schools based on the relationships that it had with the principals to ensure that the school would cooperate with data collection.

Figure 3.1. Locations of Treatment and Comparison Schools



SOURCE: Google Maps with markers added by the authors.
NOTE: Purple markers represent treatment schools, and blue markers represent comparison schools. The blue border represents the Islamabad Capital Territory border.

Samples

In this section, we describe the student, teacher, and principal samples at baseline and follow-up.

Baseline Data Collection

Baseline data collection occurred in October 2023 in the four sectors in Islamabad Capital Territory previously mentioned. At the outset of our data collection, we planned to collect data from 80 schools, 160 teachers (80 TFP Fellows and 80 comparison teachers), and 4,800 students (2,400 students in TFP schools and 2,400 students in comparison schools). Projections for student sample sizes assumed an average class size of 30 students per teacher. Table 3.4 details the achieved study sample sizes for schools, teachers, and students at baseline (separated by grade level and by treatment and comparison condition).

Table 3.4. Sample Sizes by Instrument Type at Baseline

Instrument	Total Sample	Treatment	Comparison
Surveys			
Principal surveys	80	41	39
Teacher surveys	162	82	80
Grade 4 student surveys	1,950	1,006	944
Grade 6 student surveys	1,251	793	458
Grade 7 student surveys	1,698	895	803
Student mathematics assessments			
Grade 4	1,074	525	549
Grade 6	551	294	257
Grade 7	735	381	354
Student English assessments			
Grade 4	1,104	566	538
Grade 6	455	282	173
Grade 7	550	269	281
Student science assessments			
Grade 4	410	217	193
Grade 6	478	261	217
Grade 7	654	327	327

We collected data from 41 treatment and 39 comparison schools. Teacher survey data were collected from 82 TFP Fellows and 80 comparison teachers. Student surveys were collected from an approximately equal number of students in treatment schools (n = 2,694) and comparison schools (n = 2,205), for a total of 4,899 students. Mathematics assessments were collected from 2,360 students (1,200 in treatment schools and 1,160 in comparison schools). English assessments were collected from 2,109 students (1,117 in treatment schools and 992 in comparison schools). Science assessments were collected from 1,542 students (805 in treatment schools and 737 in comparison schools).[18]

Follow-Up Data Collection

Follow-up data collection occurred in May 2024. Table 3.5 presents the sample sizes for schools, teachers, and students for the subset of study participants who participated at baseline that could be matched to data collected at follow-up. There was minimal attrition at the school level. There was some attrition at the teacher level (approximately 6 percent of the treatment group and 9 percent of

[18] Teachers taught specific subjects in the school, so we only collected an assessment from students if their TFP Fellow or matched comparison teacher was teaching the given subject.

the comparison group). Attrition at the student level was more substantial, averaging around 28 percent across all instruments. The primary reasons for student attrition from the baseline to the endline include absenteeism (13 percent of the baseline sample), student mobility (5 percent of the baseline sample), and dropout (6 percent of the baseline sample).

Table 3.5. Sample Sizes by Instrument Type at Follow-Up

Instrument	Total Sample	Treatment	Comparison
Surveys			
Principal surveys	78	40	38
Teacher surveys	150	77	73
Grade 4 student surveys	1,455	763	692
Grade 6 student surveys	913	589	324
Grade 7 student surveys	1,219	620	599
Student mathematics assessments			
Grade 4	776	402	374
Grade 6	408	217	191
Grade 7	501	275	226
Student English assessments			
Grade 4	855	427	428
Grade 6	330	206	124
Grade 7	425	187	238
Student science assessments			
Grade 4	272	147	125
Grade 6	332	188	144
Grade 7	472	224	248

Quantitative Study Student Sample

All students who could be matched from baseline to first follow-up were included in the analytic sample. Table 3.6 summarizes the background characteristics of students retained in the analytic sample. Approximately one-third of the students in the analytic sample were male, and the sample is somewhat unevenly distributed across grade levels; there were more students in grades four and seven than in grade six. The most common home language was Urdu (39 percent), although there were substantial percentages of students who spoke Punjabi (28 percent) and Pushtu (19 percent). Nearly all students had electricity inside their home (97 percent), but fewer than half had access to a computer (37 percent) or an air conditioner (19 percent). Fifty-five percent of students reported having access to the internet. About one-third (29 percent) reported being worried about not having enough food to eat in the past seven days. The average household size reported by students was about

eight people. Lastly, around 35 percent of students reported that they had repeated a grade. We examine the baseline equivalence of the treatment and comparison groups in Appendix A. Although many of the differences are statistically significant, very few of them are practically significant when compared with the 0.25 standardized mean difference (SMD) benchmark that is used in the *What Works Clearinghouse Procedures and Standards Handbook*'s quasi-experimental studies to determine whether two quasi-experimental groups are equivalent (National Center for Education Evaluation at the Institute of Education Sciences, 2022).

Table 3.6. Background Characteristics of Students, Analytic Sample

Characteristic	Percentage of Students	*n*
Gender (male)	34	1,211
Grade		
Grade 4	41	1,455
Grade 6	25	913
Grade 7	34	1,217
Home language		
Baluchi	2	54
Hindko	2	66
Kashmiri	4	128
Pothohari	2	76
Punjabi	28	1,000
Pushtu	19	628
Sindhi	1	29
Siraiki	2	78
Urdu	39	1,383
Multiple languages	Less than 1	4
Others	1	47
Home characteristics		
Electricity inside home	97	3,489
Access to internet at home	55	1,970
Television at home	73	2,620
Smartphone at home	85	3,044
Access to a computer at home	37	1,315
Dish or cable network at home	57	2,032
Generator, uninterruptible power supply, or solar at home	32	1,162
Refrigerator at home	82	2,936

Characteristic	Percentage of Students	*n*
Room cooler or air cooler at home	56	2,002
Air conditioner at home	19	695
Washing machine at home	77	2,770
Water pump at home	43	1,533
Sui gas connection at home	50	1,783
Motorcycle at home	65	2,327
Car at home	38	1,914
Food insecurity	29	1,054
Household size (mean, SD)	7.9 (4.0)	3,484
Number of boys under grade 10 living at home	1.8 (1.8)	2,726
Number of girls under grade 10 living at home	1.8 (1.8)	2,525
Repeated a grade	35	1,215

NOTE: SD = standard deviation. The analytic sample was restricted to students who participated in both the baseline and follow-up survey administrations (*N* = 3,585).

Quantitative Study Teacher Sample

All classroom teachers who could be matched from baseline to follow-up were included in the analytic sample (i.e., we removed teachers who did not participate in follow-up data collection and who joined schools at the first follow-up). Classroom teachers completed surveys that included information about their background characteristics, classroom organization, and responsibilities in the classroom (summarized in Table 3.7). About 27 percent of teachers lived in the community in which they taught. About three-quarters (71 percent) of teachers had worked at their schools for less than three years. Teachers reported spending the most time per week planning or preparing for lessons (mean of five hours) and marking or correcting student work (mean of five hours).

Table 3.7. Background Characteristics of Teachers (Analytic Sample)

Survey Item	Percentage of Teachers	*n*
Teacher characteristics		
Live in community	27	39
Worked at this school		
Less than three years	71	103
Three to five years	5	7
More than five years	25	36
Worked as a teacher		

Survey Item	Percentage of Teachers	*n*
Less than three years	47	69
Three to five years	9	13
Six to ten years	6	9
More than ten years	36	53
Technical qualification		
Bachelor's degree (B.A., B.Sc.)	16	24
Bachelor of Education	5	8
Master's degree (M.A., M.S., M.Phil.)	42	62
Master of Education	17	25
Other	18	27
Teacher responsibilities (Mean, SD)		
Planning or preparation of lessons and lesson notes (hours per week)	5 (6)	144
Marking or correcting student work (hours per week)	5 (5)	141
Counseling students (hours per week)	3 (5)	145
Participation in school management (hours per week)	2 (3)	143
General administrative work (hours per week)	2 (3)	137
Professional development activities (hours per week)	4 (6)	143
Communication and cooperation with parents or guardians (hours per week)	3 (3)	144
Engaging in extracurricular activities (hours per week)	2 (3)	136

NOTE: The analytic sample was restricted to teachers who participated in both baseline and first follow-up data collection (*N* = 150).

Quantitative Study Principal Sample

All principals at study schools completed surveys that included information about their background characteristics. Table 3.8 summarizes this information. About 33 percent of principals were male, and most principals (57 percent) had been working as principals for more than five years. About two-thirds of principals held a master's degree.

Table 3.8. Background Characteristics of Principals (Analytic Sample)

Characteristic	Percentage of Principals	*n*
Gender (Male)	33	26
Worked as a principal		
Less than three years	30	23
Three to five years	13	10
More than five years	57	43
Highest degree earned		
Master's (one year)	4	3
Master's (two year)	66	51
M.Phil.	21	16
Ph.D.	6	5
Other	3	2

NOTE: The analytic sample was restricted to schools that participated in both baseline and first follow-up data collection (*N* = 79).

Analytic Methods

In this section, we describe the analytic methods that we used to address the three quantitative research questions. Detailed information about our estimation procedures is provided in Appendix B.

Analyses of Student's Academic Achievement and Social and Emotional Learning

As described previously, our quasi-experimental design had two steps to ensure that the treatment and comparison groups were as similar as possible. In the first step, we identified comparison schools that served students who were similar to students in treatment schools and focused on the geographic locations of the schools. Our objective in identifying comparison schools using a matched pair approach was to minimize the threat of selection bias by ensuring that the treatment and comparison units were as similar as possible at baseline (i.e., prior to exposure to TFP in October 2023). In the second step, we used covariate balancing methods to account for any substantial remaining differences between the treatment and comparison units. For example, in terms of student achievement, students taught by TFP Fellows have higher baseline scores in English, mathematics, and science than their

comparison peers, and in English, these differences are large and meaningful (see Appendix A for more detail).[19]

Specifically, we used the Covariate Balancing and Weighting Web App (CoBWeb), which Markoulidakis and colleagues (2021) developed, to create weights for students in the comparison group so that, on average, they more closely resembled the treatment group. We conducted weight estimation by using baseline measures of the student outcomes (empathy; growth mindset; self-management; self-efficacy; and mathematics, English, and science test scores), socioeconomic status, food insecurity, living arrangement, home language, and gender.[20] CoBWeb simultaneously implements nine different balancing algorithms, and users are able to select the algorithm that results in optimal performance in terms of minimizing the absolute SMDs between the groups for each covariate and retaining the maximum possible effective sample size. For these analyses, we used two different algorithms: a general boosted model (GBM) for the academic outcomes and a logistic regression for the SEL outcomes. The GBM algorithm uses regression trees to fit a piecewise-constant model, and the logistic regression algorithm uses a linear combination of covariates to predict treatment status, which is operationalized as a binary indicator variable (1 if TFP and 0 otherwise [Markoulidakis et al., 2021]).[21] The GBM algorithm reduced the SMDs between the two groups to less than 0.04 standard deviations on all covariates while retaining over 90 percent of the unweighted sample. The logistic regression algorithm reduced the SMDs to less than 0.01 standard deviations.

We obtained estimates of the impact of TFP on student academic achievement and student SEL by comparing the outcomes of treatment- and comparison-group students while using covariates to control for any remaining differences at baseline between the two groups after applying balancing methods (see Appendix B). We used a canonical difference-in-differences model with two periods and a single treatment (Roth et al., 2023). We estimated TFP impact using a weighted ordinary least squares model, accounting for the clustering of students within schools in the standard error estimation (White, 1980).

Analyses of Teaching Quality and the Classroom Learning Environment

We intended for our teaching quality and classroom learning environment outcomes to measure phenomena at the classroom level, and the outcomes were derived from both student and teacher surveys. Again, the first-step matching approach was insufficient to account for all baseline differences between the TFP and comparison classrooms (i.e., differences that were larger in magnitude than 0.25 standard deviations). For example, TFP Fellows were considerably less experienced than the comparison teachers and had higher baseline ratings of rigorous expectations and clarity (see Appendix A for more detail). For this reason, we used CoBWeb to estimate balancing weights that we

[19] Consistent with recommendations in *What Works Clearinghouse Procedures and Standards Handbook*, we interpret meaningful differences in baseline equivalence to be differences between treatment and comparison groups that are greater than 0.25 standard deviations (National Center for Education Evaluation at the Institute of Education Sciences, 2022).

[20] We do not include teacher-level characteristics in the student regressions. Because TFP Fellows are less experienced than the average comparison teacher, it might be the case that the treatment effect is confounded with teacher experience, which could bias our estimates.

[21] More details about this procedure are available in Imai and Ratkovic (2014).

could use in our impact estimation for classroom-level outcomes. For these analyses, we conducted weight estimation using a baseline measure of the outcome (clarity, cognitive activation, classroom management, assessment use, self-efficacy, responsibility for learning, collaboration, challenge, control, rigor, emotional safety, and care), as well as class-average socioeconomic status, food insecurity, home language, teacher gender, and whether the teacher lived in the community.[22] We selected three different algorithms. For all outcomes other than responsibility for learning and cognitive activation, we used the Covariate Balancing Propensity Score (CBPS) algorithm. For responsibility for learning and cognitive activation, we used an entropy balancing (EB) algorithm. For the outcome of control, we used the GBM. The CBPS algorithm fits a penalized version of a logistic regression subject to a constraint that prioritizes balancing on the covariates, as well as model fit. The EB approach iteratively applies a reweighting scheme until adequate balance is achieved (Markoulidakis et al., 2021). These algorithms effectively reduced the SMDs between the two groups to less than 0.25 standard deviations for all of the baseline measures of the outcomes. However, we were unable to achieve balance in terms of teacher background; in particular, there were persistent, large differences in teacher experience even after weighting. This is not surprising, given that TFP focuses explicitly on new teachers, and all TFP Fellows in the study have either no teaching experience or one year of experience. Because of this, we interpret the estimates from our classroom-level analyses with caution.

After estimating the weights, we estimated the impact of TFP on student and teacher perceptions of teaching quality by comparing the outcomes of treated and comparison classrooms. We estimated the treatment effect using an ordinary least squares model (see Appendix B) that was balance-weighted and accounted for the clustering of classrooms in schools in the standard error estimation.

Analyses of Student, Teacher, and Principal Perceptions of School Climate

Similar to our classroom-level analyses, our initial approach—identifying comparison schools by selecting geographically proximal schools that served students who were similar to students in treatment schools—was insufficient to ensure baseline equivalence of the treatment and comparison groups (for more detail, see Appendix A). For this reason, we also used balance weighting in our impact estimation for school-level outcomes.[23] Because of the large number of variables relative to the sample size, we used a reduced set of variables in our covariate balancing (see Appendix A for details). For the school-level analyses, the covariate balancing algorithm that performed best was the CBPS approach. This algorithm reduced the SMDs between the two groups to less than 0.25 standard deviations on the baseline measures of all outcomes except for two (teacher-student relationships and school climate). Because of this, we interpret the estimates from our school-level analyses with caution.

[22] Specifically, we conducted weight estimation separately for each outcome using a baseline measure of the outcome and other covariates. The complete documentation of the variables that we used for each analysis is provided in Appendix A (Tables A.4, A.6, and A.9).

[23] Importantly, while we derived some of our school-level outcomes from the principal survey, we derived other variables from either the teacher or the student survey. For teacher and student survey–derived variables, the first step was to aggregate responses to create school-level variables.

After weighting, we obtained estimates of the impact of TFP on school climate by comparing the outcomes of treated and comparison schools. The treatment effect was estimated using a weighted ordinary least squares model (see Appendix B for more detail).

Results

In this section, we present the results from our analyses.[24] Prior to presenting the results from the regression analyses, we provide descriptive analyses of the survey scales and assessment scores at baseline. The figures we present in this section display standardized effect sizes in a dot plot; circles indicate effect size estimates, and whiskers indicate a 95-percent confidence interval. A dashed, black, vertical line indicates an effect size of zero. Where relevant, asterisks indicate statistical significance.

In Table 3.9, we present baseline descriptive information about all outcome measures. In general, the assessment scores are low. The average scores are less than 0.50, which means that respondents are typically earning less than half of the available points for each assessment. For mathematics in grade six, treatment students earned only 18 percent of the available points. Conversely, the survey scales are on the high end of the scale range (all scales have a maximum value of five).

Table 3.9. Baseline Measures and Descriptive Statistics (Analytic Sample)

	Treatment			Comparison		
Measure	**Observations**	**Mean**	**Standard Deviation**	**Observations**	**Mean**	**Standard Deviation**
Student English assessments						
Grade 4	427	0.28	0.18	428	0.22	0.14
Grade 6	206	0.37	0.14	124	0.30	0.12
Grade 7	187	0.38	0.14	238	0.31	0.16
Student mathematics assessments						
Grade 4	403	0.27	0.11	395	0.26	0.11
Grade 6	224	0.18	0.12	191	0.20	0.11
Grade 7	275	0.24	0.09	226	0.21	0.08
Student science assessments						
Grade 4	147	0.40	0.18	125	0.37	0.16
Grade 6	189	0.30	0.13	144	0.30	0.10
Grade 7	224	0.32	0.12	246	0.28	0.11
Student social-emotional outcomes						
Empathy	944	2.85	0.78	839	2.80	0.80
Growth mindset	936	3.97	0.98	719	3.76	0.96

[24] Full regression tables with coefficient estimates are available on request.

Measure	Treatment			Comparison		
	Observations	Mean	Standard Deviation	Observations	Mean	Standard Deviation
Self-management	924	3.77	0.98	820	3.73	0.92
Self-efficacy	944	4.07	0.84	723	3.99	0.84
Teaching quality						
Control	53	3.50	0.30	47	3.53	0.28
Challenge	53	3.26	0.41	44	3.37	0.42
Rigorous expectations	53	4.04	0.37	44	3.92	0.41
Emotional safety	53	3.89	0.36	44	3.87	0.44
Care	53	3.73	0.40	47	3.65	0.47
Clarity	74	3.78	0.33	72	3.56	0.39
Cognitive activation	74	3.22	0.50	72	2.75	0.43
Classroom management	74	3.37	0.53	72	3.29	0.55
Assessment use	74	3.99	0.70	72	3.91	0.66
Self-efficacy	74	3.65	0.40	72	3.40	0.46
Responsibility for learning	74	3.93	0.34	72	3.58	0.28
Collaboration	72	4.68	0.82	72	4.19	0.84
School community						
Liking for school	35	4.34	0.42	31	4.36	0.37
Safety	35	3.75	0.42	31	3.87	0.47
School connections	36	3.39	0.28	32	3.22	0.34
Teacher-student relationships	40	3.80	0.51	36	4.26	0.41
School climate	40	1.99	0.43	36	1.67	0.51
Physical safety	40	3.87	0.43	36	3.94	0.45
School environment	40	3.73	0.69	35	4.31	0.41
Teaching quality	39	3.66	0.38	35	3.78	0.23

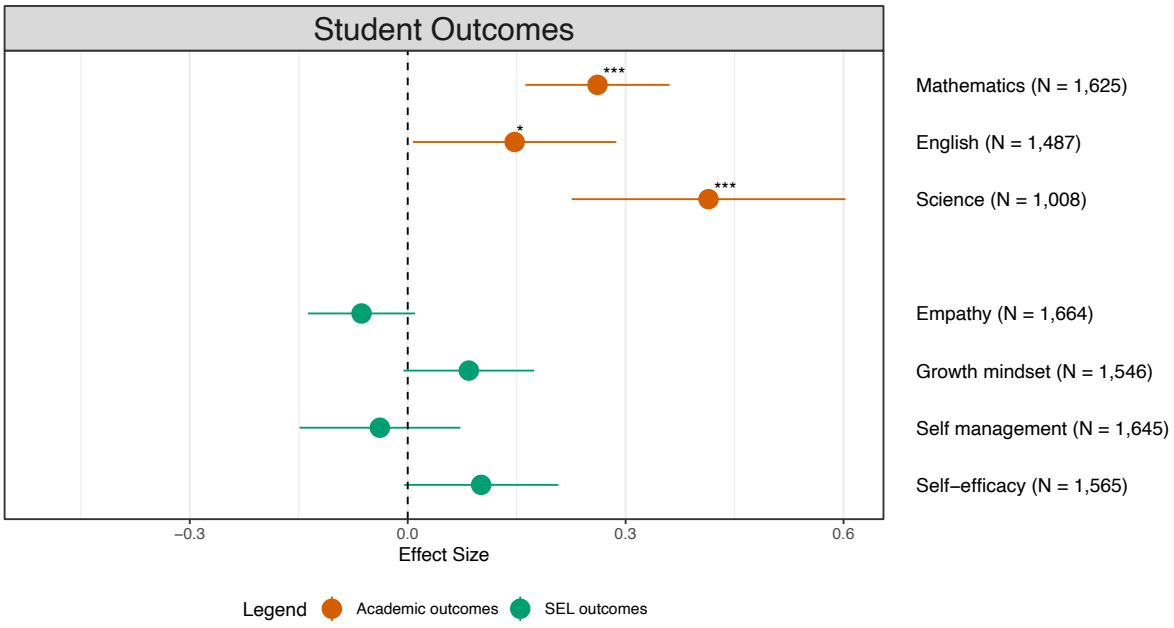
NOTE: We conducted all analyses on the entire analytic sample. Because of the block sampling and differences in subjects taught, grade level is not accounted for in weighting or regression approaches.

Impact on Whole-Child Development

Figure 3.2 shows the results of the analyses on student outcomes; orange dots represent academic outcomes. The estimates for mathematics, English, and science were positive and statistically significant at the $p < 0.05$ level. Students in classrooms taught by TFP Fellows scored 0.15 standard deviations higher in English, 0.26 standard deviations higher in mathematics, and 0.41 standard deviations higher in science than students in comparison classrooms. In terms of expected impacts on

educational outcomes, these effects can be interpreted as medium (for English) to large (for mathematics and science) effects (Kraft, 2020). The green dots in Figure 3.2 represent SEL outcomes. These estimates are generally small, not statistically significant, and no larger than 0.10 in absolute magnitude. We also tested differential effects by student gender and found a similar effect of the TFP program on whole-child outcomes for girls as compared with boys (although some of these effects were no longer statistically significant).[25]

Figure 3.2. Results of Models Estimating Effects on Whole-Child Outcomes in Teach For Pakistan Schools



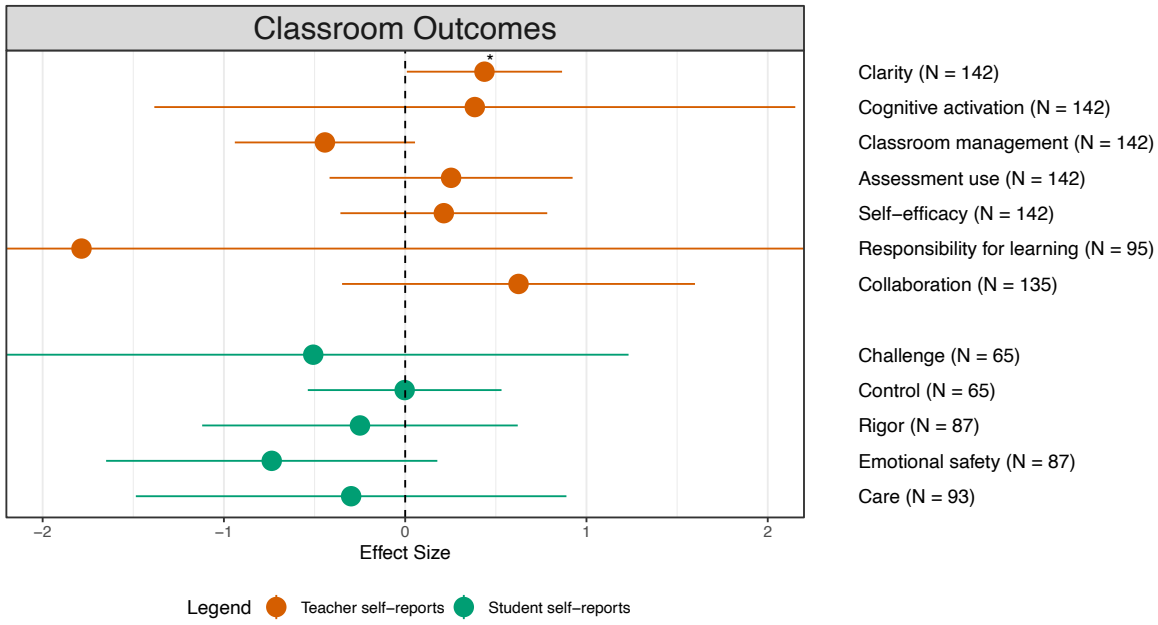
NOTE: Each dot and whisker represent a single regression. Effect sizes (dots) represent SMDs between treatment and comparison groups. Whiskers represent 95-percent confidence intervals for each estimate. All regressions are weighted by balancing weights. *** = $p < 0.001$, * = $p < 0.050$.

Impact on Perceptions of Teaching Quality

Figure 3.3 summarizes our analyses of teaching quality and the quality of the learning environment. As mentioned previously, we derived one set of these classroom-level outcomes from the student survey (represented in green) and one set from the teacher survey (represented in orange). The estimates were mixed in terms of direction: About half of the estimated effects were negative, indicating that the comparison schools scored higher than the TFP schools. However, none of these effects were statistically significant.

[25] These results are available on request.

Figure 3.3. Results of Models Estimating Effects on Teaching Outcomes in Teach For Pakistan Schools

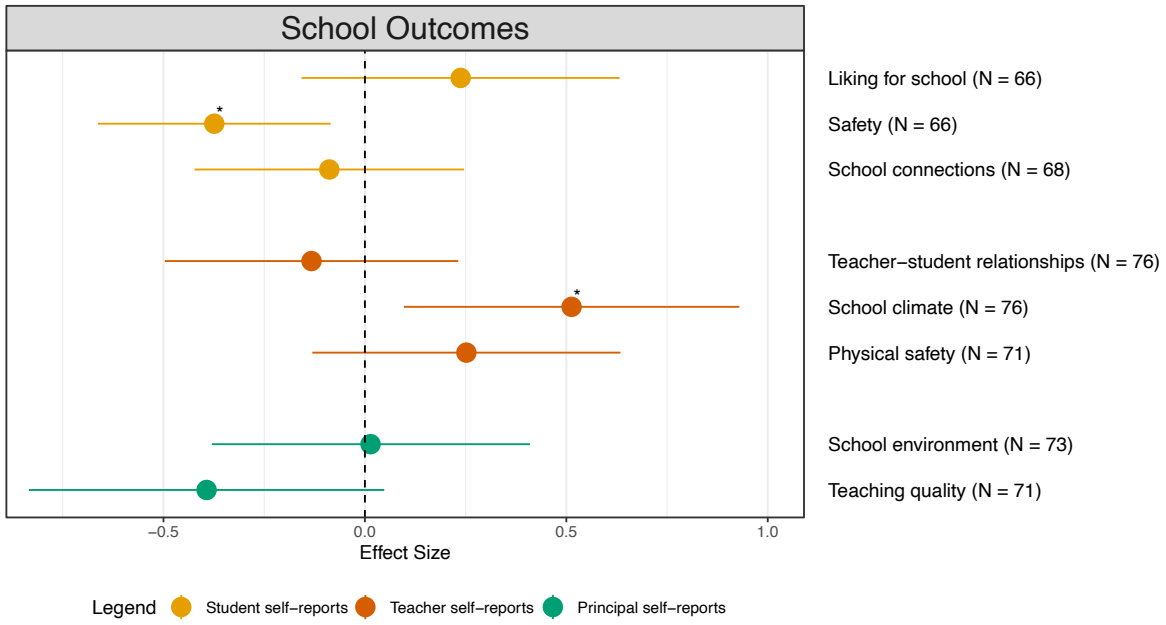


NOTE: Each dot and whisker represent a single regression. Effect sizes (dots) represent SMDs between treatment and comparison groups. Whiskers represent 95-percent confidence intervals for each estimate. All regressions are weighted by balancing weights.

Impact on School Community

Figure 3.4 summarizes the results of our analyses of aspects of school climate. We derived the outcomes in Figure 3.4 from the student, teacher, and principal surveys.

Figure 3.4. Results of Models Estimating Effects on School Community Outcomes in Teach For Pakistan Schools



NOTE: Each dot and whisker represent a single regression. Effect sizes (dots) represent SMDs between treatment and comparison groups. Whiskers represent 95-percent confidence intervals for each estimate. All regressions are weighted by balancing weights. * $p < 0.050$.

Two of these estimates were statistically significant: the effect on student perceptions of safety (which was significant and negative) and the effect on teacher perceptions of overall school climate (which was significant and positive). Overall, there were mixed effects on other school community outcomes, although these effects were not statistically significant.

Chapter 4

Qualitative Study of Teach For Pakistan

As described in Chapter 1, the qualitative study addresses questions about the contributions that various stakeholders perceive TFP Fellows make to students' whole-child development, the classroom environment, and the school community. The specific questions are as follows:

- What do various stakeholders (e.g., principals, other teachers, parents, students, and TFP Fellows themselves) perceive as the contributions of TFP Fellows on whole-child development, including student academic learning and SEL outcomes?
- How do stakeholders characterize the teaching approach and classroom environment of the TFP Fellows?
- What do stakeholders perceive as the contributions of TFP Fellows to the school community?

In this chapter, we describe the sample, data collection, analytic methods, and findings from the qualitative study.

Focal School Selection

The qualitative study included 16 TFP schools that were purposefully sampled (Creswell and Creswell, 2017; Palinkas et al., 2015; Yin, 2015) from the 41 TFP schools in the quantitative study. We determined that 16 schools would be adequate to observe notable variations and/or to confirm consistent patterns. Selection of focal schools involved input from Teach For All and TFP. The organizations had particular interest in understanding the contributions of TFP Fellows for the full two years of their fellowships because, in the second year, Fellows undertake the critical CPP. Given this, we decided that schools were eligible to participate in the qualitative study if they had at least two second-year Fellows. The requirement of at least two (rather than only one) second-year Fellows allowed respondents to reference their experiences with multiple TFP Fellows, decreasing the likelihood that perceptions were based on the idiosyncrasies of one individual. About half of the schools ($N = 21$) met these criteria.

TFP further wanted to stratify by the total number of Fellows (first- and second-year combined) in the school at the time of data collection: two versus more than two. This request was based on a hypothesis that when there is a higher concentration of Fellows in a school, they are more likely to influence school-level (and possibly student-level) outcomes. There were 11 schools with exactly two Fellows (both second-year), and the other ten schools had three to six Fellows. We randomly selected eight in each stratum. One school was subsequently replaced because it was unresponsive to our recruitment outreach. The sampling was performed with placement data from October 2023 and

April 2024. Numbers changed slightly once data collection began in May 2024. In the end, among the 16 schools in our final sample, ten had two second-year Fellows, and the remaining six had three or four second-year Fellows. One school serves only boys, eleven schools serve only girls, and four are coeducational.

Data Collection

The qualitative study used five data-collection protocols:

- in-depth interviews with TFP Fellows
- in-depth interviews with principals
- focus groups with non-TFP teachers
- focus groups with parents of students taught by the TFP Fellows
- brief interviews with students of TFP Fellows.

We developed the protocols with input from Teach For All and TFP. We referenced the TFP program's theory of change and mapped interview and focus group questions onto the key evaluation questions. Therefore, protocol questions addressed perceived cognitive and noncognitive development in students, TFP Fellows' classroom practices and environments, their contributions to the school communities, and their relationships with other educators, students, and parents. The protocols differed slightly depending on which stakeholder was interviewed, but in structure and content they were similar, particularly the four protocols with adult participants. The protocol for the brief student interviews consisted of seven main questions focused on TFP Fellows' classroom practices (e.g., classroom management, academic expectations, relating with students) and students' classroom experiences and perceptions of themselves as learners (e.g., whether they believed they were capable of learning anything, whether they came to school organized and prepared to learn, and the role their teacher played in supporting them in these respects). See Appendix B for a mapping of the protocol questions to the evaluation questions. The full interview and focus group protocols are available on request.

Prior to formal data collection, we trained RCons lead staff on qualitative data-collection methodology and the specific protocols. We also developed and shared a qualitative data-collection field manual in which we detailed procedures and guidance on topics, including the objectives of the data-collection protocols, roles and responsibilities of the data-collection team, tips for conducting interviews and focus groups, audio transcription protocols, and research ethics. Subsequently, the lead staff trained the rest of the field staff who would conduct the data collection. RCons staff translated all protocols into Urdu; TFP program leaders cross-checked and made revisions as needed.

Before formal data collection, RCons staff piloted the protocols for the TFP Fellows interview and the non-TFP teachers focus group at one nonsampled TFP school. We debriefed the experience and were aided by a native Urdu-speaking graduate student researcher who listened to the recordings and read through the notes. After our debrief, we provided suggestions to improve the data-collection process and data quality. Notably, the pilot interviews and notes were briefer than we had expected. We encouraged RCons to ask more follow-up questions to elicit more-detailed explanations and descriptions. TFP suggested that RCons consider the cultural nuances around gender and power

dynamics. Subsequently, RCons aimed to have one female staff member per pair of data collectors so that the female data collector's presence and cofacilitation might put female respondents—particularly mothers attending the parent focus groups and female students—more at ease, which might have encouraged them to speak more freely.

RCons staff approached sampled schools to schedule data collection. They obtained all necessary approvals from the school principals, recruited non-TFP teachers and parents to participate in focus groups, and secured consent from parents to interview their children. We targeted students who had been taught by the TFP Fellows between October 2023 and May 2024. In keeping with the focal grades of the quantitative study, we targeted students in grades three through eight. On the day of the data collection, RCons randomly selected among the returned consent forms to obtain the sample of four students per school to interview. Students verbally assented to the process.

RCons formally collected interview and focus group data in May 2024. RCons conducted the procedures at each school in teams of two: one lead facilitator and one notetaker and cofacilitator. As designed, the qualitative data collection at each school should have consisted of one 60- to 75-minute in-depth interview with all of the school's second-year TFP Fellows (N = 32–64 total), one 60-minute interview with the principal (N = 16 total), one 60-minute focus group with two or three non-TFP teachers who had worked with the TFP Fellow(s) or knew them well (N = 32–48 total), one 60-minute focus group with three to six parents of students taught by TFP Fellows (N = 48–96 total), and 15-minute brief individual interviews with four students taught by TFP Fellows (N = 64 total). Actual data collection met these targets. Across the 16 sampled schools, we spoke with 36 TFP Fellows, 15 principals, 42 non-TFP teachers, 80 parents, and 64 students. All interviews and focus groups were conducted in Urdu. RCons recorded audio from interviews and focus groups with participants' permission. Tables 4.1 and 4.2 present the data-collection summary and the interview and focus group sample, respectively.

Table 4.1. Qualitative Study Data-Collection Summary

Informant	Procedure	Expected Number of Informants per School (Total Across 16 Schools)	Actual Number of Informants per School (Total Across 16 Schools)	Actual Average Length (Minutes)
TFP Fellows	Interview	2–4 (32–64)	1–3 (36)	79
Principals	Interview	1 (16)	0–1 (15)	45
Non-TFP teachers	Focus group	2–3 (32–48)	2–4 (42)	44
Parents	Focus group	3–6 (48–96)	3–11 (80)	41
Students	Brief interview	4 (64)	4 (64)	22

Table 4.2. Qualitative Study Interview and Focus Group Sample

School	Second-Year TFP Fellows (Grade Taught Second Year)[a]	Principal (Years of Experience)	Non-TFP Teachers (Average Years of Experience)	Number of Parents	Student Grades
1	Three males (8–9)	Male (17)	Three males (19)	7	6–8
2	Two females (8)	Female (10)	Two females (16)	5	6–8
3	Two females (8)	[Not interviewed]	Two females (9)	3	6–8
4	One female (9–10)	Female (16)	Three females (10)	3	6–8
5	Two males (3–4)	Male (1)	Three males (10)	4	5
6	Two females (3–5)	Female (1)	Three females (22)	11	3–5
7	Two females (3–5)	Female (5)	Two females (29)	6	3–5
8	Two females (8)	Female (1)	Three females (10)	4	6–8
9	Three females (4)	Female (14)	Two females (12)	5	3–6
10	Three females (6–8)	Female (1)	Three females (12)	4	7–8
11	Two females (3–5)	Female (2)	Three females (14)	7	3–5
12	Three females (3–5)	[Not provided]	Two females (14)	5	3–5
13	Two females (4–5)	Female (8)	Two females (23)	4	4–5
14	Two females (3–6)	Female (5)	Four females (17)	4	4–6
15	Two females (7–8)	Female (4)	Three females (13)	4	6–8
16	Three females (6–8)	Female (1)	Two females (14)	4	6–8

[a] This column refers to the grades that these TFP Fellows taught in May 2024.

Data Analysis

RCons's bilingual team produced notes and transcripts in English of each interview and focus group, following a transcription protocol. We performed quality assurance, spot-checking audio recordings and notes and working iteratively with RCons to ensure accurate and comprehensive notes. We imported the final dataset into Dedoose (a cloud-based application that supports mixed-methods research) for coding and analysis.

First, we performed coarse-grained coding, largely to categorize the excerpts into the key questions of interest for the qualitative study. For example, under the top-level code "Evaluation Question 1: School Community," we had two main subcodes that indicated TFP Fellows' "Main Contributions to School Community" and "Relationships." Under the subcode "Relationships," we had broad codes for "Relationship with principal," "Relationship with other teachers," "Relationship with parents," and "Relationship with students." Subsequently, we performed second-level coding to surface key themes within each main question of interest for the qualitative study. We generated an initial set of thematic codes using anticipated responses to the interview and focus group questions. We also allowed for

emergent codes. Following established qualitative research practices (Lincoln and Guba, 1985; Lincoln and Denzin, 2003; Miles, Huberman, and Saldaña, 2014), we generated a codebook and defined and refined codes as needed during the coding process. Furthermore, the qualitative research team met to discuss coding progress, address emergent issues, and resolve ambiguities. See Appendix B for a mapping of the main evaluation questions to the interview and focus group protocol questions, to relevant constructs from the TFP theory of change, and to our coding scheme.

After coding, we performed thematic analysis (Guest, MacQueen, and Namey, 2011). We drew on established techniques, such as looking for repetition, similarities, and differences among data (e.g., similar responses across respondents in different roles), to identify themes (Ryan and Bernard, 2003; Bernard, Wutich, and Ryan, 2016). We took steps to ensure the integrity of our findings, such as examining both confirming and disconfirming evidence (Denzin, 2006). We conducted two collaborative sensemaking sessions with Teach For All and TFP partners as we coded and began to draw emergent themes from the data. The goals of these sessions were to discuss the findings and emergent themes; identify gaps, nuances, or details that might be missing from the emergent themes we had already identified; and reevaluate the data with a different perspective. We summarized key thematic findings in a matrix in which the rows were the focal schools and the columns were the guiding evaluation questions. The matrix facilitated cross-site analyses of findings across the 16 schools.

Findings

In this section, we first present findings related to stakeholders' perceptions of TFP Fellows' contributions to students' academic and social and emotional development. Second, we describe how stakeholders characterized TFP Fellows' teaching approach and classroom environment. Third, we present themes related to TFP Fellows' contributions to the school community. The themes we present are based on cross-site analyses; generally, our findings are consistent across the 16 focal schools. Where relevant and notable, we mention contrary or outlying findings. In text boxes, we highlight the voices of students describing their experiences in TFP Fellows' classrooms.

We remind readers of several limitations. First, because the focal schools are a purposeful sample that we identified in collaboration with TFP, we cannot be sure that the qualitative study findings generalize to all TFP schools. For the same reason and because the two studies address different questions and somewhat different constructs, we caution against trying to reconcile the qualitative findings with those from the quantitative study. Third, because principals, non-TFP teachers, and parents are not present in the TFP Fellows' classrooms on a daily basis, they likely have less exposure to and knowledge about what TFP Fellows' teaching practices and student outcomes look like. If a stakeholder did not mention a certain pedagogical practice or outcome, it does not necessarily mean that the Fellow did not address it but rather that the stakeholder's perspective might have been constrained.

Whole-Child Development

In the two subsections that follow, we describe key findings related to stakeholders' perceptions of Fellows' contributions to two aspects of whole-child development: students' academic or cognitive outcomes and social and emotional outcomes.

Academic Outcomes: English, Mathematics, and Critical Thinking

Across the 16 schools, stakeholders perceived that TFP Fellows affected students' academic learning. Although some interviewees and focus group participants spoke of progress in general literacy, including reading and speaking Urdu, many of them highlighted English as a specific area of notable improvement. Educators, parents, and students themselves remarked on students' ability to speak English with greater fluency and confidence. One TFP Fellow provided the following assessment of their students: "In the start, they did not understand English, but I kept talking to them in English. . . . They will start understanding it one day. Now they are not only able to understand but reply to me in English as well." Students also developed vocabulary, spelling, and grammar skills to enable them to read stories and write responses in English. TFP Fellows at different schools described that students who previously did not know simple English words were able to write in sentences and paragraphs after the TFP Fellows' teaching, and students who were previously able to write sentences could now produce essays and stories. The principal at another school found students' ability to write in English "impressive," saying, "I . . . checked their English paper [exam] . . . and I am really very happy with them."

Student Voices on Academic Development

"I couldn't speak English, so [the TFP Fellow] asked to speak only English in class. Through this, I got motivated and can now speak a bit."

"I faced difficulty in math and English. Now, [the TFP Fellow] helped me, and my English and math have improved a lot."

"[The TFP Fellow] wants us to score good marks. She says to understand the concept and then learn it; don't just rote memorize."

Stakeholders also perceived student growth in mathematics and science; some claimed that some students performing below their grade levels were now performing at their grade levels. At one school, non-TFP teachers, parents, and TFP Fellows all noted large improvements in mathematics, with one Fellow saying, "I have seen incredible growth in numeracy. . . . When I came, [the students] were at grade two level, although they were in grade seven. The growth level has been three levels in this year." Parents and other interviewees at other schools relayed similar anecdotes, for example that, after TFP Fellows' teaching over the course of about one school year, children who previously struggled with basic arithmetic could now solve multiplication and division problems independently. Additionally, some stakeholders across schools spoke of students initially finding the subject of science abstract, difficult, and unapproachable but then growing to understand and become interested in the concepts, in part because of TFP Fellows' use of engaging hands-on activities and project-based approaches.

However, stakeholders' perceptions of academic progress were not unanimous across or within schools. One principal was particularly disappointed: "We have never seen any positive change in kids. Our board results this year [are] the worst ever. [TFP Fellows] were given three subjects—English, science, and mathematics. Kids failed in these subjects." Moreover, non-TFP teachers at one school thought that TFP Fellows increased students' self-confidence toward learning but did not positively influence academic results. According to these teachers and a few others at a small number of schools, the mathematics results for students of TFP Fellows were reportedly negative.

Notably, stakeholders perceived that students' academic growth was linked to their development of conceptual understanding and critical thinking and analysis. At nearly all schools, stakeholders

remarked on a shift from passive, rote learning (or cramming) to active thinking. A related common pattern was that students initially did not ask questions in lessons or know how to think outside the box. With TFP Fellows' teaching and guidance, students became more willing to pose questions and think critically. One TFP Fellow reported, "In the science lecture, [students] asked wondrous questions, for example about the Big Bang theory . . . . They think about these things themselves rather than relying on me to tell them." And one principal summed up students' cognitive development as follows: "Due to the TFP Fellows, the academic records of the kids have improved. Their critical thinking [has] also developed a lot. TFP Fellows guide the children in such a way that kids think critically now. They ask the questions. In this way, their minds have been polished."

Social and Emotional Outcomes: Grit and Confidence, Collective Responsibility, and Agency

One noncognitive outcome that multiple stakeholders across schools said they observed in TFP Fellows' students is grit, or resilience in the face of challenges. Stakeholders, including parents, characterized students who used to give up easily as now choosing to struggle or persist through difficulty or set a goal to achieve next time. One TFP Fellow described an experience with students who received below-passing grades on an assignment. The students approached the Fellow to say that they would double their scores on the next task. Students have also been observed discussing questions they got wrong with peers to try to better understand the concepts. One TFP Fellow described that students developed "this level of perseverance that, even if they know that they cannot do something, they still try it."

Student Voices on Grit and Confidence

"I firmly believe in myself. If one believes in herself, she can do anything in the world. She can progress. She can go to the moon. She can do anything."

"I believe that if I work hard, I will learn anything. If something seems to be difficult, I learn it by repeating it again and again."

"I can learn anything. The teacher has given me the confidence. First, I used to think that either I can do it or not, but now, if I make a mistake, the teacher explains. Therefore, now I think that I can learn everything. There is nothing impossible; you have to work a little harder."

Multiple stakeholders spoke of qualities adjacent to grit, specifically that they perceived that TFP Fellows' students have developed a stronger sense of confidence, responsibility, and self-efficacy toward learning. Stakeholders saw these qualities manifesting in students who have become more confident communicators and courageous advocates who ask for support. One parent said, "The confidence level of the children has improved. They don't hesitate to express their point of view." Other interviewees spoke of students who did not wait for teachers to tell them an answer but took the initiative and responsibility of finding answers themselves. Students have even asked for time to work on mathematics problems without teacher guidance or modeling because they wanted to solve the problems themselves. Some students also reportedly track their own assignment scores and performance as a way of taking responsibility for their own progress. One parent summarized the change they observed as follows: "The children have become very confident now. Now, they keep doing the task until they get it completed. Also, they have become responsible. . . . There is no need to ask the children again and again to do the task; now, they do it themselves."

Consistently across the focal schools, stakeholders perceived greater collaborative spirit in TFP Fellows' students and a sense of responsibility for one another's learning and actions. This corresponds to TFP's desired outcome of collective responsibility. One manifestation of collective responsibility is that students seemed to have fewer conflicts and became more considerate of one another. One TFP Fellow noted that "in the beginning, [students] were not very friendly and had competition with each other." Others noted that students argued or "had an aggressive tone while interacting with each other and [they] used to fight." Over time, stakeholders perceived that students evolved to speak more kindly to each other, with one TFP Fellow saying, "They are more empathetic to each other." A few interviewees described experiences observing students who were communicating productively with each other to find solutions to conflicts.

Another aspect of collective responsibility that TFP Fellows, non-TFP teachers, principals, and students at various schools noticed was that students took ownership of the groups they were working in and helped ensure that everyone in the groups completed their tasks. One TFP Fellow described what they observed: "[Students] have developed collective responsibility. . . . If someone is having a rough day, their [classmates] collaborate with [them]. If one student has an achievement, all of them celebrate it together. If a student fails, they think of it as an opportunity to grow." Students appeared to develop a mindset that they could extend support to others, particularly new or struggling students who would benefit from extra help. One non-TFP teacher described this as students "promoting team spirit. . . . This is very good, as you become sincere for your team, and you know that you have to take the one with you who is relatively weak."

Stakeholders that we interviewed further noticed that students taught by TFP Fellows had become more aware of their rights and had developed a sense of agency. In this respect, students not only used their confidence and voices in academic matters and took on responsibility for their peers' learning, but they also began using their voices to advocate for their collective rights. Students have shown this agency on several important issues. One is their right to an education. One TFP Fellow described that if a teacher did not come to class (or was unpunctual), students would approach the teacher and tell them to "please take our class," showing that the students have become "cognizant of the marginalization they face" and have evolved to "ask for whatever they think should be done directly." In this way, students have become more active in fighting for their future opportunities.

Student Voices on Sense of Collective Responsibility

"Earlier, students were used to quarrelling with one another and were not on speaking terms. . . . Now, after the arrival of TFP teachers, students don't quarrel with each other. We work together. Help the student who needs it. If there would be [a] fight among students, the teacher resolves it. She guides us [so] that we all have to learn together."

"Before [the TFP Fellow], the students did not get along with each other and only focused on their studies, but when [the TFP Fellow] came, she told us that we should help each other in studies as well. The students who are better in studies help the weak students."

"We share our lunch with others. If someone is sick, we ask her about her sickness. We care about others and behave nicely with each other."

"If a student is going through some trouble, we all ask her the problem. If she resists to tell everyone, we move aside and let her close friend ask her. And then we look for solutions together."

Another issue on which students have shown agency is corporal punishment. TFP Fellows noted that students learned that "they cannot be emotionally or physically harmed by anyone." Stakeholders have observed students standing up for a supportive learning environment, for example, by bringing a teacher's use of corporal punishment to the principal's attention or even voicing their disagreement to the teacher directly. Students have also organized petitions or requests, for example, writing a letter to the principal, signed by the entire class, to have a broken fan in the classroom repaired or to request cleaner washrooms. One principal spoke positively of the changes he observed in students' confidence and agency as follows: "I see a visible change in [students'] behavior. We Pakistanis think that submissiveness means disciplined. Discipline is something else. [TFP Fellows' students] style of talking and movement have changed. I saw a big difference in the students as compared with other students. I want other students [to] get the same opportunity." Many stakeholders shared this perspective. However, some non-TFP teachers and one principal interpreted students' increased activism and confidence in expression as disrespect or arrogance, and one teacher commented that "the children got out of hands from us."

Teaching Approach and Classroom Environment

In the previous section, we described stakeholders' perceptions of TFP Fellows' contributions to aspects of whole-child development—the student outcomes. Here, we take one step back in the theory of change (Figure 2.1) and describe TFP Fellows' pedagogical and relational practices with students in the classroom. Specifically, we present four main themes that, taken altogether, suggest a student-centered approach. The themes are teaching for mastery and conceptual understanding, motivating students to learn, cultivating a caring classroom environment, and empowering students to lead. In some instances, stakeholders directly linked a practice to student outcomes. We summarize their perceptions but caution that we cannot establish causal links with the data we have collected.

Teaching for Mastery and Conceptual Understanding

Across all 16 schools in our sample, principals, teachers, students, and TFP Fellows themselves highlighted teaching from students' current level and teaching for mastery as key characteristics of TFP Fellows' instructional approach. Fellows described administering diagnostic assessments and discovering that many students' knowledge and skills were about three grade levels below their assigned grades. Drawing on their TFP training, the Fellows knew that they could not just teach the required syllabus; instead, they worked out a long-term plan to address the learning gap and help students achieve mastery. In many cases, *mastery* is defined as grades of about 80 to 90 percent. Rather than moving on after teaching a lesson or after students had attempted an assignment, TFP Fellows required students to demonstrate their clear understanding of the targeted learning content or independent use of the skill. Fellows reported that they identified student learning objectives and provided additional support until the students had met the objectives. This meant that the TFP

Fellows often retaught lessons, provided remedial support, gave feedback on assignments, and had students redo tasks to ensure that they had sufficiently learned the topic or skills to proceed to the next unit or level. Many students appreciated the additional time and repeated practice that the Fellows provided. Several principals shared this sentiment, and one said, "[The TFP Fellows] guided the students well. They gave individual attention to the weak students who were not performing well." In this respect, TFP Fellows held every student to the same high expectations and expected them to work hard to meet the goals.

Student Voices on How Teach For Pakistan Fellows Taught for Mastery

"If we find anything difficult, [the TFP Fellow] explains it again and again until we learn it."

"If a student is unable to finish the assigned work, [the TFP Fellow] doesn't punish her but asks her to try to finish it again."

TFP Fellows appeared to have drawn on a repertoire of less traditional methods of pedagogy and engagement in teaching for mastery. For example, contrary to emphasizing memorization and rote learning, TFP Fellows focused on teaching for conceptual understanding. One Fellow said that they focused on "conceptual clarity" and teaching for depth of student learning rather than breadth. To support this type of learning, TFP Fellows encouraged curiosity in students and used modern pedagogical practices that were largely unused in Pakistani schools. That is, all TFP Fellows, along with most principals, parents, and non-TFP teachers who participated in our interviews, noted that Fellows used visual aids, hosted practical demonstrations, made the learning relevant to student or community needs, and engaged students in interactive, hands-on lessons and project-based learning. One TFP Fellow described this strategy: "First, our focus is to develop a hook and develop the interest of students in the topic by asking questions or showing them a model. . . . We ask what they know about it. [Our] objective is to make them understand the concept." Another TFP Fellow contrasted their practice with traditional instructional methods, saying, "When we first came here, the students did not ask questions, even if they did not understand the lesson. . . . We made them good enough at speaking that it develops curiosity in them to ask questions and try to look for deeper knowledge of the topic and develop mastery in the subject. . . . First, the students just memorized what they were taught. They did not think of creative ways they could learn more about it. . . . Through our . . . practices, they were able to get more understanding and achieve mastery." Not everyone regarded the activity-based approach as effective. Non-TFP teachers at one school countered that TFP Fellows' approach does not support students to do well on state assessments: "Students learn the concepts, but they are unable to write it in exams. So, they don't show the required results. TFP Fellows focus on activities more and not on the written tests. . . . [Their students] don't memorize, so they are unable to get good marks in exams."

Finally, TFP Fellows reportedly promoted mastery by monitoring student progress, encouraging students to track their own progress, and using such data to inform instructional next steps and differentiate instruction for students as needed. One Fellow described this practice, saying, "If we saw that a particular student gets [a] low grade continuously, then we worked with them by going to their level. If there is a group activity, and I give a question, I see which student is unable to do it so [I can] work with that student. . . . I used data to teach him on a level he is able to understand."

Motivating Students to Learn

Besides enacting instructional practices to support student learning, stakeholders noted that TFP Fellows also engaged in a variety of practices that activated students' motivation to learn, perseverance through challenges, and growth mindset. Some TFP Fellows did this through explicit lessons and examples. One Fellow described such a lesson: "We help [students] understand: What is perseverance, and what is grit? We let them see visually one man carrying a stone trying to climb up the mountain and falling down. He tries again, and, in the mid-way again, he fell down. But at the end, he succeeded. [Then] we let students see the progress tracker. . . . You are at 50 [percent]. You need to achieve 80. This method developed grit in the students." Other TFP Fellows highlighted cases in which students persevered and significantly improved their grades; Fellows shared these students' accomplishments to help other students realize that "if a person does not give up and has grit, perseverance, and resilience, he can accomplish anything." TFP Fellows also used praise, words of encouragement, incentives, and rewards to motivate students.

Student Voices on How Teach For Pakistan Fellows Motivated Students

"[The TFP Fellow] has high expectations of us. She wants us to become something big. She insists we participate in all the activities of the class. She appreciates and encourages us."

"[The TFP Fellow] motivates us, saying that whatever you practice or work hard at, you will achieve. There is no difficult task. We only need to do work hard and show perseverance."

"[The TFP Fellow] always motivates us, saying that if we study exhibiting full interest, then we may achieve our goals. . . . They let us write our life goals. And push us to do our best efforts to achieve them."

Cultivating a Caring Classroom Environment

In general, according to various stakeholders' perceptions, TFP Fellows cultivated a classroom community characterized by care and collective responsibility rather than conflict and aggression. Stakeholders reported that TFP Fellows built a foundation of mutual trust and respect with students by consistently showing warmth and affection toward them. This is in stark contrast to the common practice of meting out corporal punishment to correct student behavior. In return, students felt that they could approach the TFP Fellows with any issue or concern. One non-TFP teacher characterized the relationship TFP Fellows had with students as follows: "[The Fellows] treat the students with love and affection. They get along with the students. . . . Their class atmosphere is so loving that we also want to join their class. The students who don't talk with us easily talk to them." Although most stakeholders praised TFP Fellows for this relational approach to teaching, one principal remarked that "[the TFP Fellows] methodology is very advanced, but their classroom discipline is very weak."

TFP Fellows also frequently used collaborative learning approaches to foster teamwork and a sense of responsibility for one another's learning. Fellows reportedly grouped students by mixed ability to encourage students to help each other and to provide students with leadership opportunities within teams. They also designed activities such that students could not accomplish the tasks on their own; completing an activity required peer support and interactions. Another strategy some TFP Fellows employed was minimizing the use of the singular pronoun *I* in the classroom and instead using the collective pronoun *we*. In one classroom, the motto is "We can," implying that students ought to work together as one. One Fellow described that, as a result of this approach, "[students] stand together for certain issues in [the] classroom. They have unity in [the] classroom, unlike in the beginning, when they had a grudge with each other." TFP Fellows also encouraged students to compliment and show appreciation or empathy toward their peers. When conflict did arise, Fellows helped students to resolve tensions and coached them on ways to avoid or prevent future conflict. Despite the TFP Fellows' efforts, though, in a small handful of schools, stakeholders reported that there were students who struggled to get along with their classmates and continued to engage in or be subjected to bullying.

Student Voices on How Teach For Pakistan Fellows Cultivated a Caring Classroom Environment

"Even if someone does a very small, good thing, [the TFP Fellow] asks everyone to clap for her. [The TFP Fellow] acts as if something really big has happened even if it is a very small deal."

"Our teacher tells us if anyone's bag is torn, don't make fun of him. If he is poor, don't make fun of him. Anyone can have bad times. Help as much as you can."

"Our regular teachers teach the children with a harsh behavior and sometimes also hit and scold them, but the TFP teachers teach us with love and affection. They give us equality, so all students work better . . . here, everyone is treated the same by the TFP [Fellows]."

"I comfortably got [the TFP Fellow's] help. Some teachers are short tempered. We don't go to them, but our [TFP Fellow] is friendly—we easily go to them."

Empowering Students to Lead

Our interviewees noticed that TFP Fellows supported students to develop beyond academics and relationally; they empowered students to be agents of their own learning and future. Fellows have supported students to become more critically self-aware in terms of their areas of strength and improvement. Furthermore, they have helped students develop a vision for themselves and work toward those goals. In particular, TFP Fellows have helped students become aware of academic and career options and have helped inspire and grow ambition among students. One principal testified to this, saying, "The vision of the students for the next five years is clear. Now every kid thinks about his future. The kids have become more conscious about the field or profession they want to adopt in the future." One TFP Fellow raised the issue of marginalization and inequity facing students. They discussed with students the differences between government-funded schools and colleges and higher-quality private schools and colleges and explained how students can attend better schools and colleges. This Fellow helped students understand scholarship opportunities and even took students on a trip to Islamabad, where most of the universities are located. As one TFP Fellow observed, this and similar approaches have helped students "have dreams of themselves."

Fellows also empowered students by providing them with opportunities to voice opinions and make decisions in the classroom. For example, students helped establish classroom rules and decide activities. One TFP Fellow reported sometimes giving students the autonomy to decide instructional methods: "The students feel safe and can tell us if they are tired and will not be able to focus on the class. Then we would change our way of teaching, which makes them feel that they have ownership of what they are doing. We have a feedback system in our class. The students once told me how the lesson could be better so . . . they can better understand the topic."

TFP Fellows also expressly supported students in understanding and asserting their rights and helping make their school communities and world a better place. In particular, TFP Fellows consciously raised the topic of corporal punishment and explicitly taught students not to accept it: "[The students] got to know that the punishments in the school that had been there for many years were . . . wrong. They felt that it was because of their own deeds, but that was wrong. They didn't know that it was their right that they shouldn't be punished physically. So, they started speaking for their rights." TFP Fellows also described students engaging in activities that reflected their growing sense of agency in the larger community. For example, students distributed flyers in their community to raise awareness about the environment. Empowering students seems to be an intentional focus of TFP Fellows. In the words of one Fellow, "Agency is the most important goal of our class. We created opportunities for [students] to show their agency. We made them understand their rights and what is their role in the school. They thought that their role was to come and listen in the school rather than talking. We made them understand that they are a stakeholder and they can show agency and how they can utilize it . . . . We ask for their opinion about the matters of the country and how they would solve it."

TFP Fellows' efforts to motivate and empower students might be related to Fellows' own shifting mindsets and perspectives on inequities in the education system over the course of the fellowship, specifically their growing realization that students' low achievement is not the fault of the students but rather stems from failures in the system and students' backgrounds, which students have no control over. For example, one Fellow said, "I thought that the students were not willing to study, but, after coming, I saw passion in them and their efforts and realized that if I put my confidence in them and believe in them, they can do anything . . . . The inequity of learning . . . [stems from students being from] a disadvantaged community."

Others extended this type of thinking to parents and other educators, recognizing that they face challenges as well. For example, some Fellows had thought that parents did not prioritize education but then realized that "differences in resources . . . don't let them prioritize it." For example, parents might be unavailable for meetings because they have work

Student Voices on How Teach For Pakistan Fellows Empowered Students

"First, we didn't know how we should prepare properly before coming to school, but [the TFP Fellow] has guided us. Now, it remains in our mind even at home for what purpose we are going to school, and what we have to learn, and what are today's tasks."

"I was not the same in the beginning. . . . Before ending the class, [the TFP Fellow] explains to us the activities of the next day. Now, even if she doesn't tell us, we know how and what to prepare for the next day."

"[The TFP Fellow] taught us problem solving. Once, the fans of our classroom were not working properly. We wrote an application to the principal, and she got them repaired."

obligations: "They are facing economic crises. That is why they could not get time [to come to the meeting]." As for fellow educators, one TFP Fellow said, "I thought that if the school is not performing well overall, then it is because of the teachers and the principal. But, after coming here, I have realized that it is not the problem of the school but is part of a greater problem. For example, [students] do not have the writing desks, which affects their writing." Fellows came to regard educators and parents as partners in efforts to improve conditions for students, with one concluding, "Everyone is playing their role to bring change."

Contributions to the School Community

In this section, we describe key findings related to the contributions that stakeholders perceived TFP Fellows made to the school community—contributions that the TFP program would like to see endure beyond the Fellows' placement in the schools. We discuss TFP Fellows' intentional engagement of parents as partners in the education of the students, commitment to ending corporal punishment and fostering a student-centered approach to classroom management, and additional contributions via the CPP.

Engaging Parents as Partners

According to interviewees in all 16 schools, TFP Fellows cultivated supportive relationships with parents. All second-year TFP Fellows identified a school-wide need to improve parental engagement and worked to improve it as part of their CPPs. Typically, teachers did not engage parents in their children's learning. However, as part of their training, TFP Fellows were encouraged to bridge the gap between school and home. Rather than viewing parents as figures removed from their children's learning or as an obstacle, TFP Fellows regarded parents as collaborators in growing the children in cognitive and noncognitive ways.

Principals, other teachers, and parents themselves reported that TFP Fellows actively engaged parents via in-school meetings and home visits. By these accounts, TFP Fellows called and messaged parents on cell phones from a few times per week to a few times per month to apprise parents of their children's academic progress. In these calls, TFP Fellows encouraged parental involvement in the children's education. TFP Fellows suggested that parents pay closer attention to their children and their work, helped parents identify their children's areas of struggle and ways to improve, and educated parents on the use of technology in their children's learning. Parents expressed appreciation for TFP Fellows guiding them to better understand how to support their children's education and social and emotional development at home. One parent said, "If a child has any problem, [the TFP Fellow] calls us and [informs us] about that issue. They guide us well about the children."

TFP Fellows' efforts to involve parents had a perceived impact beyond the direct relationships they built with parents. By connecting with parents and inviting them into the school building, TFP Fellows helped instill a sense of responsibility in parents and acculturate them to the ways in which they could engage with educators to support their children's academic learning and overall development. One principal remarked on how TFP Fellows helped parents grow a sense of awareness and ownership of their role as the drivers of their children's education: "[The TFP Fellows] developed a sense of responsibility in parents. . . . Parents think that it is only the responsibility of teachers to

teach. But these [Fellows] develop a sense of awareness in parents, too." In this way, some parents have begun to view not only the TFP Fellows but the school as a whole as approachable. One parent described this shift, saying, "There is visible change in the school climate after [the TFP Fellows'] arrival. Earlier, parents were not asked to visit the school. Now, parents come regularly."

Championing Student-Centered Approaches to Classroom Management

According to our interviewees, TFP Fellows built collegial and collaborative relationships with most other teachers and the principal, and they leveraged such relationships in their efforts to shift other educators' beliefs about and practices of using corporal punishment as the default way to manage student behavior—a topic addressed in the TFP program. Fellows gained the trust of other teachers and their principals by being responsive to colleagues' requests, demonstrating competency in teaching their classes, and regularly sharing materials, resources, and pedagogical practices. For example, TFP Fellows communicated the needs and progress of their students to their principals and helped their principals in various tasks, such as analyzing student assessment data and developing class schedules. TFP Fellows also aided other teachers with lesson planning, providing classroom management strategies, and hosting professional development workshops on technology and other topics.

Having gained such trust, many TFP Fellows broached the topic of discipline directly in conversations with other teachers and their principals. Stakeholders reported that TFP Fellows explained the importance and impact of teaching with love and encouragement rather than with harsh words and actions. TFP Fellows also explained that, with less use of corporal discipline and more student-centered approaches to classroom management, students were more likely to build stronger relationships with educators and feel more positively about their education. TFP Fellows provided their colleagues with alternatives that they could use to motivate students to listen to them and respect them without using corporal punishment. Reportedly, as a result of such efforts on the part of TFP Fellows, other teachers began adopting more-constructive classroom management strategies, such as using a rewards chart and giving stars to students who behaved appropriately. One Fellow reported that the entire school implemented their classroom management strategies, and the school saw a "drastic decrease in corporal punishment." A parent at another school perceived such changes, saying, "We see change in the school due to these [TFP Fellows]. Following these [Fellows], other teachers don't punish the students, too; they teach them with care." TFP Fellows expressed with pride and hope the contributions that they believed they had made to the school community. One Fellow said, "We have brought love in this school." Another said, "Now other teachers have also realized how to develop a relationship with the students so that they can continue to work on the children after we are gone from here."

Additional Contributions to the School Community

Through their CPPs, TFP Fellows attempted to make additional contributions to their school communities that were intended to have a lasting impact beyond their time in the school. TFP Fellows' projects varied in focus, spanning students' academic and well-being needs. At several schools, TFP Fellows identified a need to improve students' literacy outcomes and decided that providing better access to reading materials could help achieve this goal. As a result, they built libraries in classrooms or in the school, placing shelves and stocking them with collections of books. They engaged

in fundraising efforts and collected donations to realize this project. Through CPPs, TFP Fellows also aimed to increase students' career awareness. One Fellow reported learning that only two students who had graduated from their school had ever gone on to university, which prompted the reflection that students lacked career counseling and exposure to future possibilities: "When we first came here, the students only opted for three choices: doctor, teacher, and joining the army. They thought that there were [only] three professions which they could join." TFP Fellows invited external speakers to their schools to provide exposure to a variety of career options, and they carried out career counseling sessions during which they imparted to students that "if they are good at something, they can pursue it as a career." Other Fellows focused on students' need for physical activity to address whole-child well-being. Projects of this kind included making a play area on school grounds and/or acquiring equipment for games, such as table tennis and cricket. TFP Fellows at other schools addressed students' access to basic needs in school. These included clean water, working water fountains, and female students' access to menstrual hygiene products.

At several schools, TFP Fellows' CPPs focused on building teacher capacity, specifically expanding their colleagues' facility with technologies. TFP Fellows taught their colleagues basic uses of computers and common applications, such as those available through Google. A Fellow described this as "digitally empowering the teachers. Our teachers are not familiar with surfing the internet or [how] to google, so we have tried to make them feel empowered digitally rather than feeling left out." A Fellow at a different school reported that, after being exposed to these supports, one of his colleagues began recording student grades in Microsoft Excel sheets and used them to calculate grade point averages. This saved the teacher considerable time compared with writing down such data and making manual calculations; moreover, this method allowed the principal to access the data.

Discussion

We intended for the qualitative study to provide descriptive and narrative examples of how TFP Fellows interacted with students and the school community and their contributions to student outcomes. We cannot make causal claims about the link between TFP Fellows' teaching approaches and the student outcomes, nor can we attribute TFP Fellows' practices to the TFP program. We did find considerable consistency in practices and themes across the sampled schools, which might signal that many aspects of the TFP fellowship program's theory of change are manifesting. Moreover, sometimes, TFP Fellows specifically identified how aspects of TFP programming supported them to make their reported contributions. They and other stakeholders also attributed some aspects of students' development to their classroom practices. In the rest of this section, we briefly and cautiously consider these potential linkages for the sample of schools in the qualitative study.

Overall, there appear to be notable similarities between the topics addressed during TFP Fellows' training and coaching (for an overview, see Chapter 2) and their pedagogical and relational approaches. In terms of the foundations of teaching, the TFP program focuses on differentiating instruction for students, teaching at the students' level, and using data to inform next steps. TFP Fellows' mastery-based learning approach reflects such principles. Fellows' student-centered approach to motivating students and empowering them to be agents of their own learning and future seems to follow from the TFP program's emphasis on cocreating goals with students and helping them "invest

in visions of their own selves." Moreover, the TFP program championed classroom environments characterized by love and mutual support. According to interviewees, TFP Fellows not only taught students with care and patience but also helped students get along with one another and minimize conflicts through the use of collaborative learning and positive reinforcement strategies. In these and other ways, the TFP program seems to have influenced the teaching approaches of TFP Fellows. Much more directly, the TFP program's CPP component supported Fellows to engage in collective and system leadership and to mobilize other stakeholders toward collective action with school- or community-level impacts.

It is conceivable that, in the sampled schools, the TFP Fellows' teaching approaches helped foster targeted skills in many students, particularly grit, collective responsibility, and agency. For example, mastery-based learning requires students to attempt assignments again and again. With TFP Fellows' warm support and with encouragement from peers, students might have developed what stakeholders perceived as determination and confidence to persist through academic and other challenges. Likewise, TFP Fellows providing students with choices, helping them imagine possible futures, and teaching them about their rights might have contributed to students' emerging sense of agency.

Although most stakeholders' characterizations of the TFP Fellows' teaching approaches and contributions aligned, there were some differences in their assessments of these approaches and contributions. For example, there was general agreement that TFP Fellows used more student-centered instructional approaches: They avoided rote learning methods, encouraged inquiry and discussion, and used hands-on activities. The Fellows also managed student behavior without the use of corporal punishment. At one school, however, the principal and non-TFP teachers thought that the TFP Fellows did not use enough discipline; they perceived the classroom as not well managed and students as disrespectful and "too confident" because they questioned authority. For the most part in this chapter, we have aimed to present what stakeholders reportedly observed regarding TFP Fellows' practices and perceived student outcomes without their appraisals of such practices. We recognize that cultural and social beliefs and perspectives around what it means to be a *good* or *well-taught* student or a *well-behaved* class will differ. Stakeholders who subscribe to a more traditional view of teaching and education might believe that a well-behaved class is a silent one with obedient children. At one school, the teachers viewed the TFP Fellows as having "no class control" with "kids mak[ing] noise in their classes." A more modern view might champion students being dynamically engaged in group projects or discussions. In this way, stakeholders' appraisals could be more of a reflection of their backgrounds and belief systems than an observation of the contributions of the TFP Fellows or program.

Chapter 5

Summary and Recommendations

In this report, we have examined the extent to which the TFP leadership development program supports whole-child development, improves teaching quality and the learning environment, and changes perceptions of the school community. The analyses in this report provide multiple perspectives on TFP in the Islamabad Capital Territory during the data collection period between October 2023 and May 2024. Using data from student mathematics, science, and English assessments and surveys administered to students, we examined evidence of the impact of TFP Fellows on students' academic achievement and social and emotional development, such as empathy, growth mindset, self-management, and self-efficacy. Using data from interviews and focus groups, we sought to elaborate on how other stakeholders—principals, other teachers, parents, and students—perceived the contributions of TFP Fellows on students' academic learning and social and emotional outcomes. We used data from student surveys, teacher surveys, focus groups, and interviews to examine the impacts that TFP Fellows had on various aspects of teaching quality and the learning environment, such as rigorous expectations and emotional safety. We used data from student, teacher, and principal surveys to characterize perceptions of the school climate and sought to understand how other teachers, parents, and students perceived the contribution of TFP Fellows to their school communities.

In this chapter, we summarize our findings from the quantitative and qualitative studies, combining the results from both studies. We organize this discussion around the three main focal areas: whole-child development, teaching quality, and school community. As we noted previously, we expected there to be differences across the two studies because of the differences in the way we measured constructs and the samples used for data collection. Although the quantitative study provides systematic evidence using rigorous measures and the larger sample of 80 schools, the qualitative study provides rich evidence driven by the respondents in 16 purposefully selected schools and presents additional context and practices that can inform the quantitative findings.

We conclude with recommendations for the TFP program and similar teacher training programs.

Summary of Findings on Whole-Child Development

To examine the effect of the TFP program on whole-child development, we used surveys to collect measures of achievement and SEL from Pakistani students. As demonstrated in our quantitative study analysis, the TFP program was successful in improving student achievement in mathematics, science, and English during a school year at a greater rate than for students in comparison schools, and these effects were medium to large in size. Stakeholders who participated in focus groups and interviews emphasized noticing improvements in English and reading competency, along with improvements in mathematics and science learning, with a special emphasis on conceptual understanding and critical

thinking. The findings in both the quantitative and qualitative analyses, therefore, show strong positive effects of the program on achievement, which is one of the primary goals of the TFP program.

The quantitative study did not find evidence that student SEL improved more for students of TFP Fellows than for students of comparison teachers. TFP, in collaboration with RAND, selected three SEL scales for the study because of their alignment with the TFP theory of change. The findings from the quantitative study of the impact of the program on these outcomes were mixed, with some positive and some negative results; the effects were small in magnitude, and none reached conventional levels of statistical significance. Qualitatively, interviewees perceived that students of TFP Fellows had improved grit, confidence, and collective responsibility. Moreover, according to evidence from the qualitative study, students had a stronger collaborative spirit and demonstrated more agency. Encouragingly, many of the SEL constructs identified in the qualitative study are ones that are included in the TFP theory of change.

Summary of Findings on Perceptions of Teaching Quality

We also considered the effect of the TFP program on teaching quality and the learning environment. The scales used to measure perceptions of teaching quality in the quantitative study included student-reported measures (control, challenge, rigorous expectations, emotional safety, and care) and teacher-reported measures (clarity, cognitive activation, classroom management, assessment use, self-efficacy, responsibility for learning, and collaboration). Results from the quantitative study did not point to large or statistically significant effects on these measures from the students' or teachers' perspectives, and there were some positive and some negative effects.

From the perceptions of principals, TFP Fellows, non-TFP teachers, parents, and students in a small sample of TFP schools, we found evidence that TFP Fellows taught for mastery and conceptual understanding, motivated students to learn, cultivated a caring classroom environment, and empowered students to lead. Interviewees reported that TFP Fellows focused on addressing learning gaps and tailoring the curriculum to meet students' needs. The qualitative study, therefore, also found support for teaching practices identified in the theory of change.

Summary of Findings on School Community

One of the goals of this evaluation was to examine whether there were any indications of spillover effects, specifically ones for which we could attribute changes in the school community to TFP Fellows. Although spillover effects are not directly in the TFP program theory of change, we wanted to understand whether the perceptions of students, teachers, and principals about school climate differed across the treatment and comparison schools. In the quantitative study, we did not find any evidence of effects on school climate. We examined student-reported measures (liking for school, safety, and school connections), teacher-reported measures (teacher-student relationships, school climate, and physical safety), and principal-reported measures (school environment and teaching quality). We did not find differences in these measures in schools with TFP Fellows compared with schools without TFP Fellows. The findings were mixed, with some positive and some negative effects,

and most effects were not statistically significant. This is unsurprising given that school-level effects are not a part of the TFP theory of change.

From the qualitative study, the stakeholders whom we interviewed perceived that TFP Fellows made some school-level contributions, particularly in the areas of fostering parental engagement, ending corporal punishment, and shifting toward a student-centered approach to classroom management. Parental engagement was one of the key aspects of school community highlighted in the TFP theory of change.

Recommendations

Using the findings from the quantitative and qualitative studies, we have identified several takeaways and corresponding recommendations for the program.

The positive effects that we found on student achievement combined with evidence from the qualitative study suggest that the **TFP program should continue to encourage its teachers to engage in key practices**, such as fostering students' sense of self (e.g., grit and confidence), teaching students to collaborate, empowering them to lead, and providing supports to ensure that every student learns. TFP Fellows were reported to have developed trust with non-TFP teacher colleagues and principals, and these other educators seemed open to the student-centered approaches that TFP Fellows bring and are sharing with the school community. At least in the sample of TFP schools involved in the qualitative study, many stakeholders—students, parents, other teachers, and principals—perceived positive contributions of TFP Fellows on students' whole-child development and attributed such growth to TFP Fellows' practices and training.

At the same time, the impacts of the TFP program on the specific SEL outcomes used in the quantitative study were small and not statistically significant. **TFP should consider implementing targeted SEL interventions and curricula with explicit instruction on SEL to improve these skills.** TFP can use recent research that reviews SEL interventions and curricula and finds rigorous evidence about which programs improve specific SEL skills (Grant et al., 2017; Cipriano et al., 2023). TFP might wish to pilot-test the intervention to ensure that the materials are appropriate for the Pakistani context.

Given that there was some resistance to TFP Fellows' approaches by some principals and non-TFP teachers who subscribe to a more traditional view of teaching and education, **TFP might consider providing supports (e.g., programming, contextualized coaching) to help TFP Fellows overcome resistance from other school staff to their teaching practices.** Examples of possible topics to include in these supports are entering spaces with humility, approaching others with different orientations, navigating different value systems, and understanding assumptions about classroom environments or pedagogies. The program might also consider providing similar training to principals and non-TFP teachers in schools in which TFP operates to ensure that the successful approaches are not hampered by school administrations.

In the qualitative study, we sought out multiple stakeholders to triangulate the perceptions of TFP Fellows. **We believe that it is important for TFP to also use multiple modes to measure student outcomes and obtain a comprehensive understanding of student progress.** Although surveys provide an opportunity for students to share their perspectives, self-reporting on SEL

measures has known limitations, specifically that it can be difficult to separate the real effect of a program from bias in students' own perceptions. An alternative way to include students in data collection is to consider performance-based measures, such as SELweb, which is an online "assessment system designed to measure social thinking skills and peer relationships" (EdInstruments, undated). Another alternative is to include external, third-party observations of student behavior to provide an objective perspective.

Appendix A

Baseline Equivalence

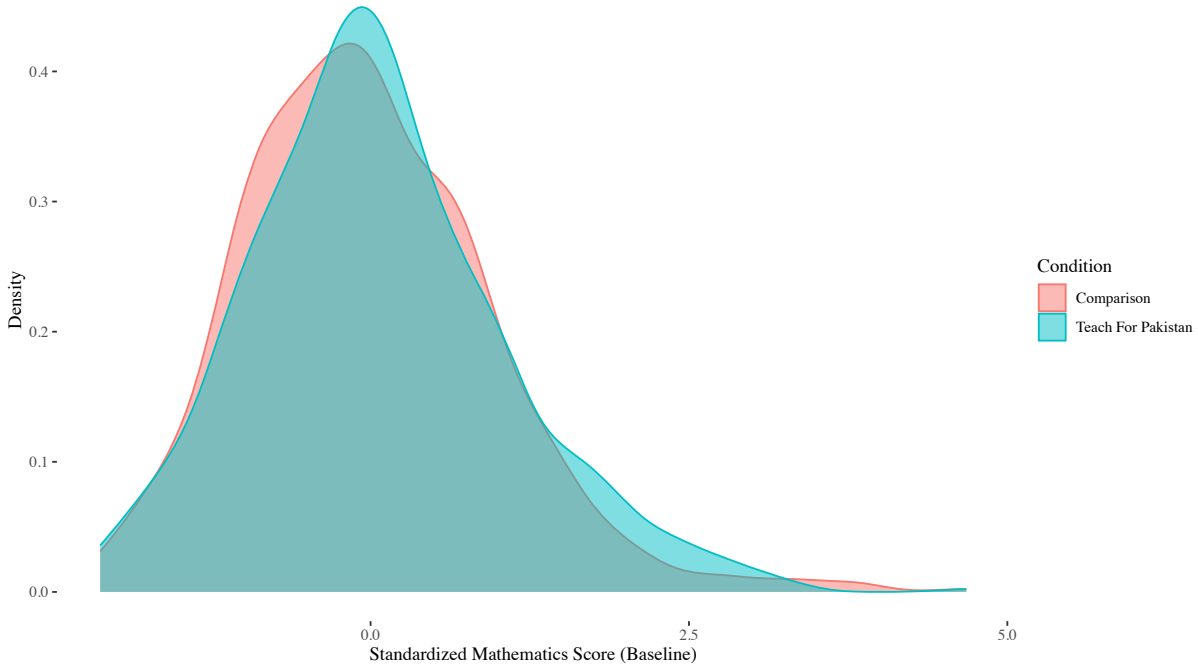
In this appendix, we provide details on the baseline equivalence of the treatment (TFP) and comparison (non-TFP) units. First, we describe the equivalence of the students. Then, we turn to the baseline equivalence of the classrooms. Finally, we present details on the baseline equivalence of the schools. In each section, we provide density plots showing common support on each baseline measure of the outcome variables and tables showing the SMDs across the quasi-experimental groups. Common support is subjectively assessed by visual inspection of the density plots. Baseline equivalence is assessed using benchmarks articulated in the *What Works Clearinghouse Procedures and Standards Handbook* for quasi-experimental studies to determine whether two quasi-experimental groups are equivalent (National Center for Education Evaluation at the Institute of Education Sciences, 2022). Specifically, we interpret a SMD smaller than 0.25 as evidence that the two groups are similar on that variable at baseline.

Baseline Equivalence of Students

Common Support

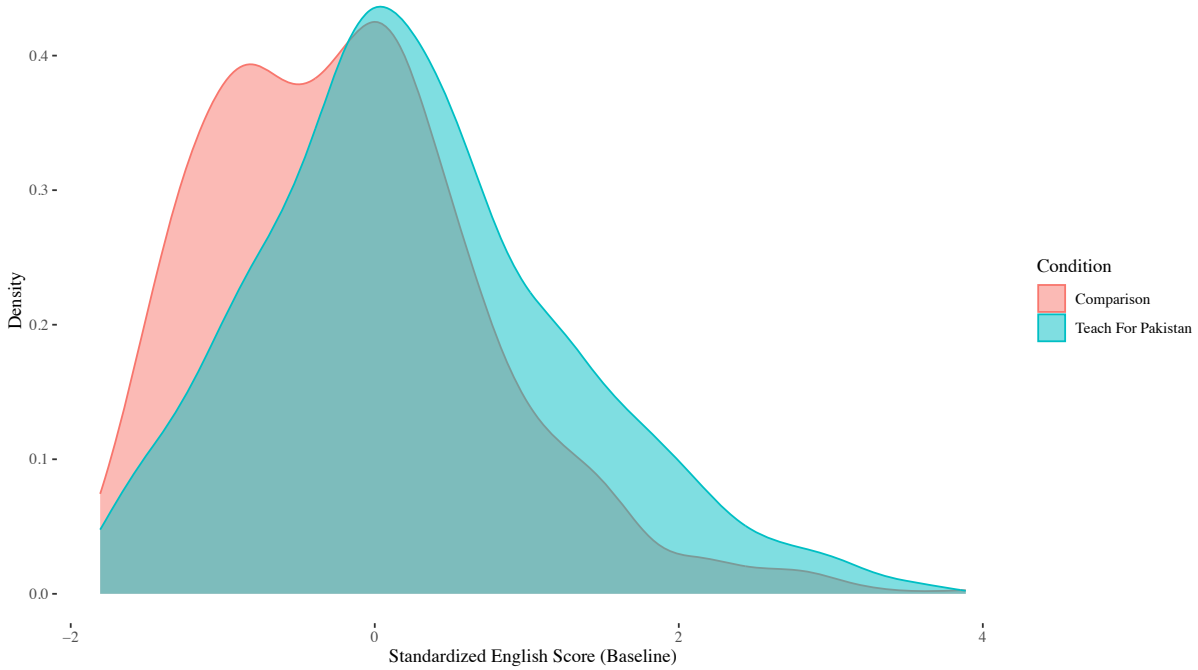
Figures A.1 through A.7 display density plots for the baseline measures of our outcome variables separately by quasi-experimental condition. The TFP students are displayed in teal, and the comparison students are displayed in pink. For each of these variables, there is considerable overlap in the distribution, providing strong evidence of common support.

Figure A.1. Density Plot of Mathematics Scores



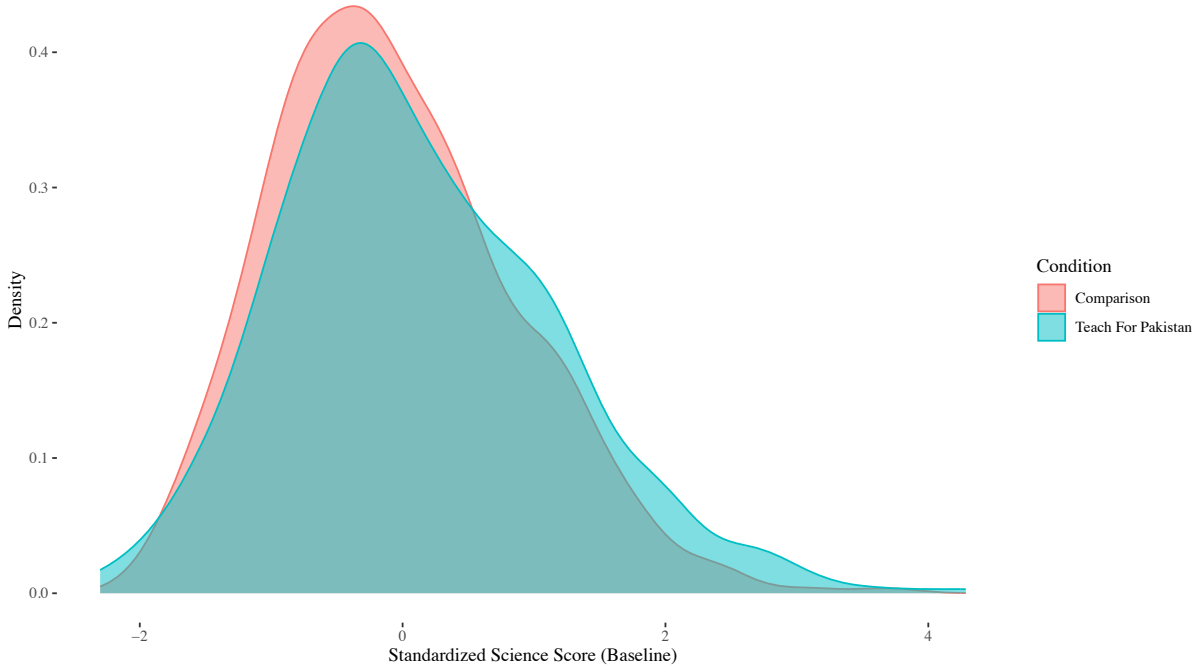
NOTE: We restricted our sample to students who participated in both baseline and follow-up data collections.

Figure A.2. Density Plot of English Scores



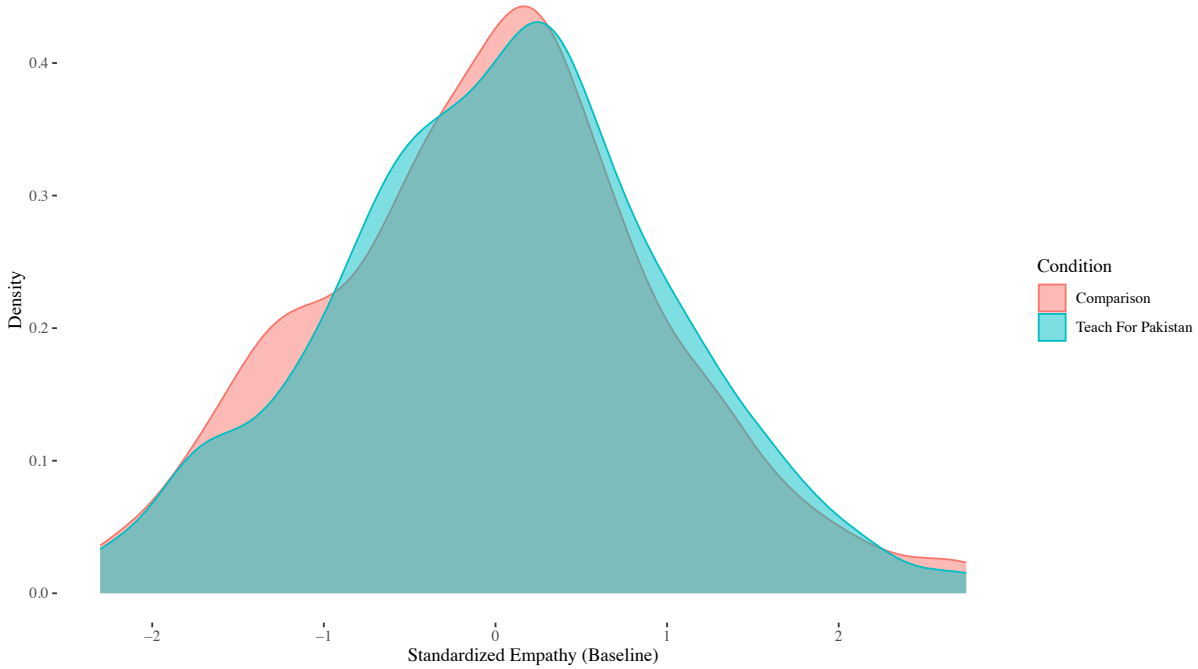
NOTE: We restricted our sample to students who participated in both baseline and follow-up data collections.

Figure A.3. Density Plot of Science Scores



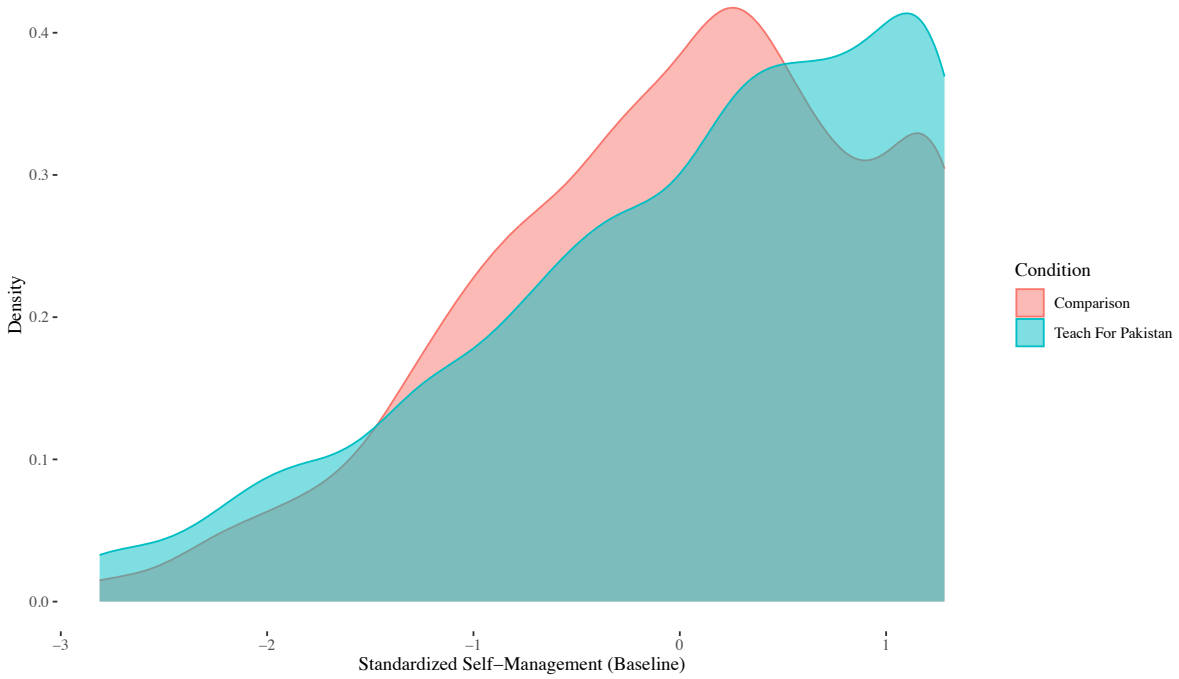
NOTE: We restricted our sample to students who participated in both baseline and follow-up data collections.

Figure A.4. Density Plot of Empathy Scores



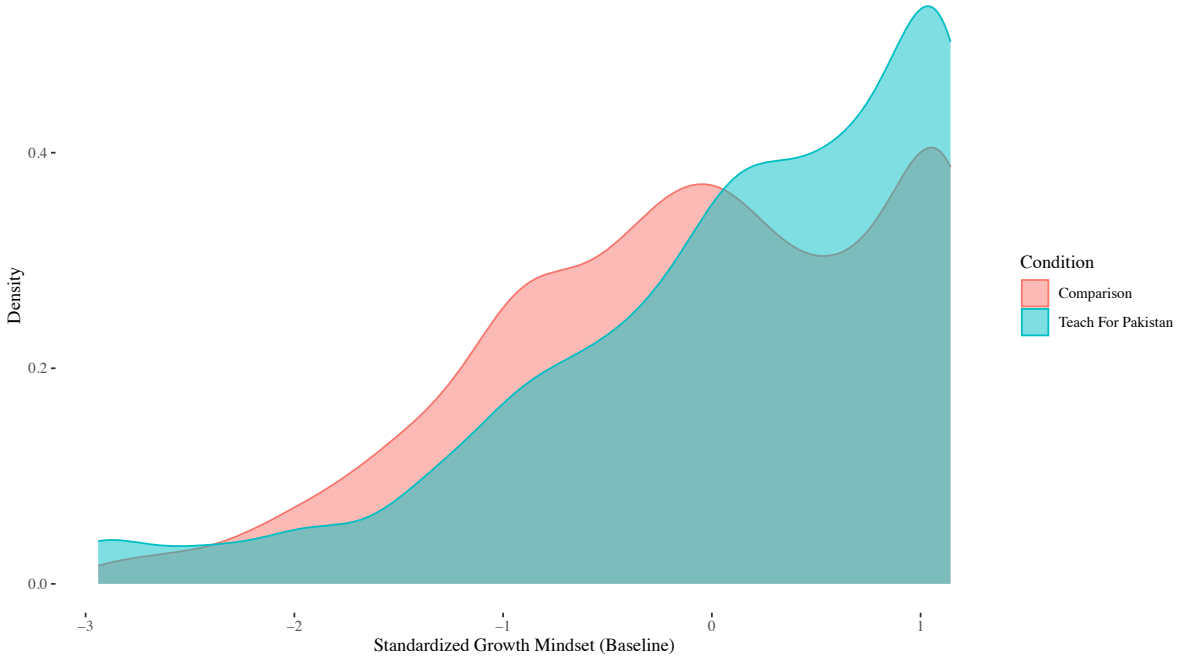
NOTE: We restricted our sample to students who participated in both baseline and follow-up data collections.

Figure A.5. Density Plot of Self-Management Scores



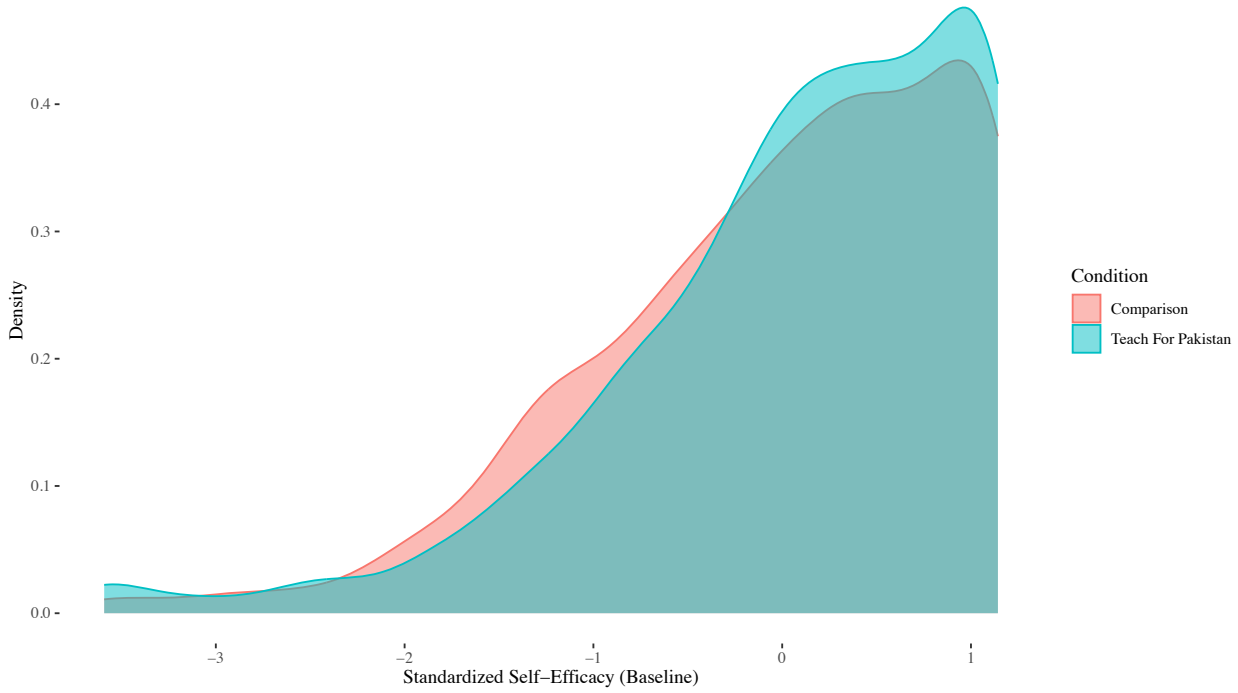
NOTE: We restricted our sample to students who participated in both baseline and follow-up data collections.

Figure A.6. Density Plot of Growth Mindset Scores



NOTE: We restricted our sample to students who participated in both baseline and follow-up data collections.

Figure A.7. Density Plot of Self-Efficacy Scores



NOTE: We restricted our sample to students who participated in both baseline and follow-up data collections.

Standardized Mean Differences of the Unweighted Sample

In this section, we examine the mean characteristics of the treatment and comparison students from the matched analytic sample and report the SMD for each of the characteristics that are included in our analytic models. As shown in Table A.1, most of the differences reported fall well below the 0.25 SMD that is used by the *What Works Clearinghouse Procedures and Standards Handbook* in quasi-experimental studies. However, there are a few notable areas in which the differences between the TFP students and the comparison students are more pronounced; in particular, scores on the English assessment were considerably higher for the TFP students. Because of the magnitude of these differences, we employed covariate balancing methods in our analyses of student-level outcomes.

Table A.1. Baseline Equivalence of Student Background Characteristics (Unweighted Analysis Sample)

Statistic	Treatment Mean (SD)	Comparison Mean (SD)	SMD
SEL skills			
Empathy	0.03 (0.99)	−0.03 (1.01)	0.06
Growth mindset	0.09 (1.00)	−0.12 (0.98)	0.22
Self-management	0.02 (1.05)	−0.02 (0.94)	0.04
Self-efficacy	0.04 (1.00)	−0.05 (1.00)	0.10
Academic performance			
English	0.20 (1.03)	−0.21 (0.92)	0.42
Mathematics	0.04 (1.02)	−0.04 (0.97)	0.08
Science	0.09 (1.06)	−0.10 (0.92)	0.19
Student characteristics			
Socioeconomic status	0.62 (0.22)	0.62 (0.21)	0.02
Food insecurity	0.33 (0.47)	0.28 (0.45)	0.12
Number of children in household	2.58 (2.81)	2.67 (2.55)	−0.03
Repeated a grade	0.36 (0.48)	0.35 (0.48)	0.02
Home language			0.18
Baluchi	0.02	0.02	
Hindko	0.02	0.02	
Kashmiri	0.04	0.03	
Pothohari	0.02	0.02	
Punjabi	0.28	0.29	
Pushtu	0.22	0.17	
Sindhi	0.01	0.01	
Siraiki	0.02	0.02	
Urdu	0.36	0.42	
Other	0.02	0.01	
Gender (male)	0.33 (0.47)	0.35 (0.48)	−0.03

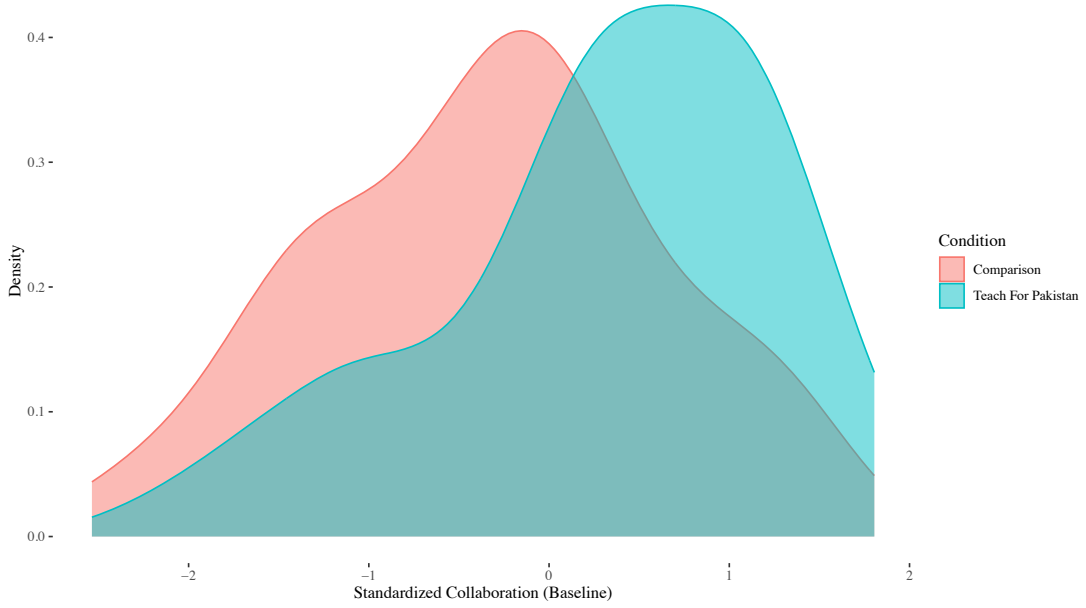
NOTE: The socioeconomic status variable is a composite constructed from home possessions (percentage of home possessions a respondent indicates having). SEL skills and academic performance variables are standardized. For categorical variables, a multivariate Mahalanobis distance represents the SMD, following Yang and Dalton (2012). We restricted our sample to students who participated in both baseline and follow-up data collections. Total N = 3,585. N_{TFP} = 1,972. $N_{COMPARISON}$ = 1,613.

Baseline Equivalence of Teachers and Classrooms

Common Support

Figures A.8 through A.19 display density plots for the baseline measures of our outcome variables separately by quasi-experimental condition. The TFP classrooms are displayed in teal, and the comparison classrooms are displayed in pink. For each of these variables, there is considerable overlap in the distribution, which provides strong evidence of common support.

Figure A.8. Density Plot of Collaboration

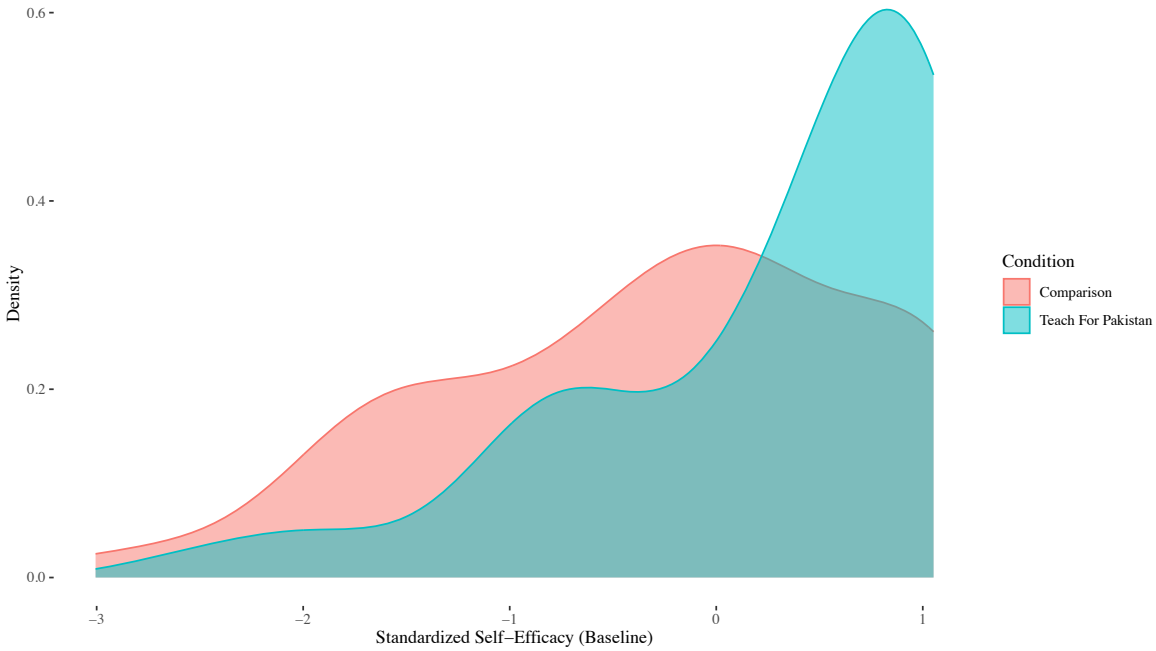


NOTE: We restricted our sample to teachers who participated in both baseline and follow-up data collections. We aggregated variables to the classroom level for analysis.

Figure A.9. Density Plot of Responsibility for Learning

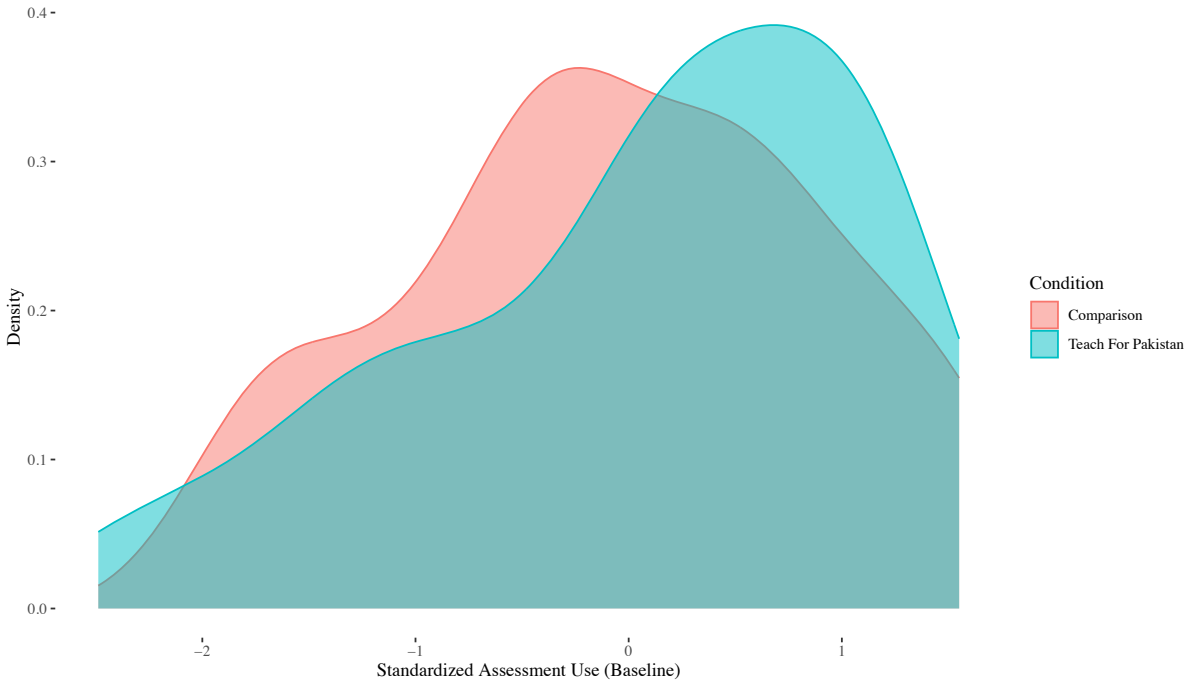
NOTE: We restricted our sample to teachers who participated in both baseline and follow-up data collections.

Figure A.10. Density Plot of Self-Efficacy



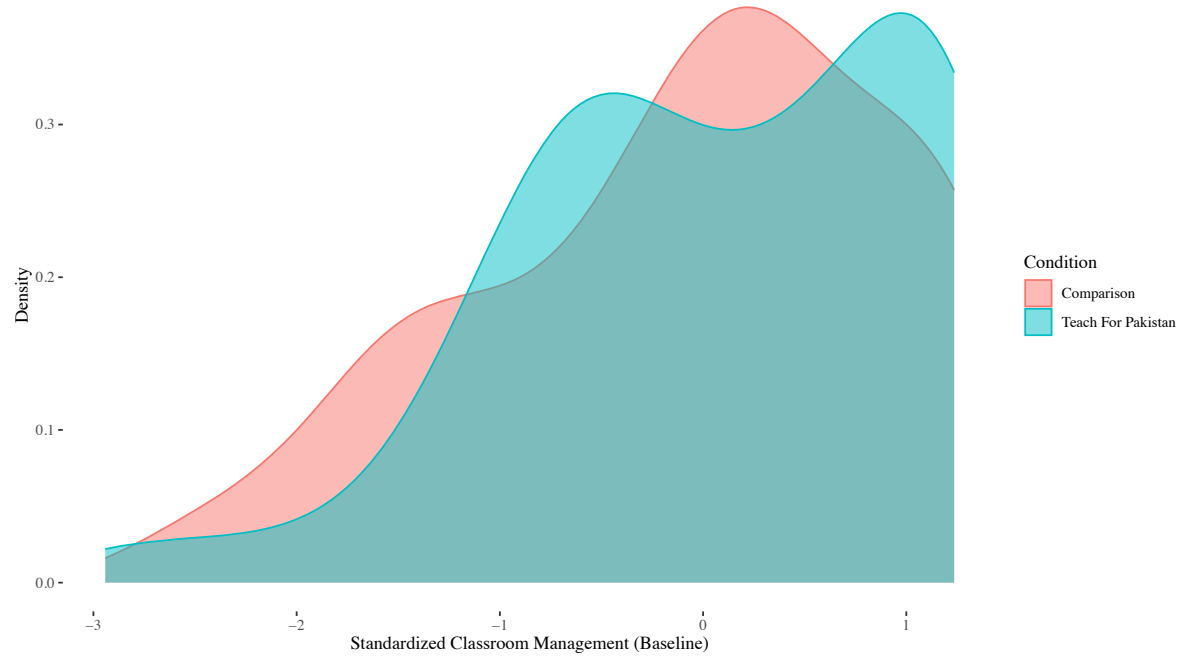
NOTE: We restricted our sample to teachers who participated in both baseline and follow-up data collections.

Figure A.11. Density Plot of Assessment Use



NOTE: We restricted our sample to teachers who participated in both baseline and follow-up data collections.

Figure A.12. Density Plot of Classroom Management

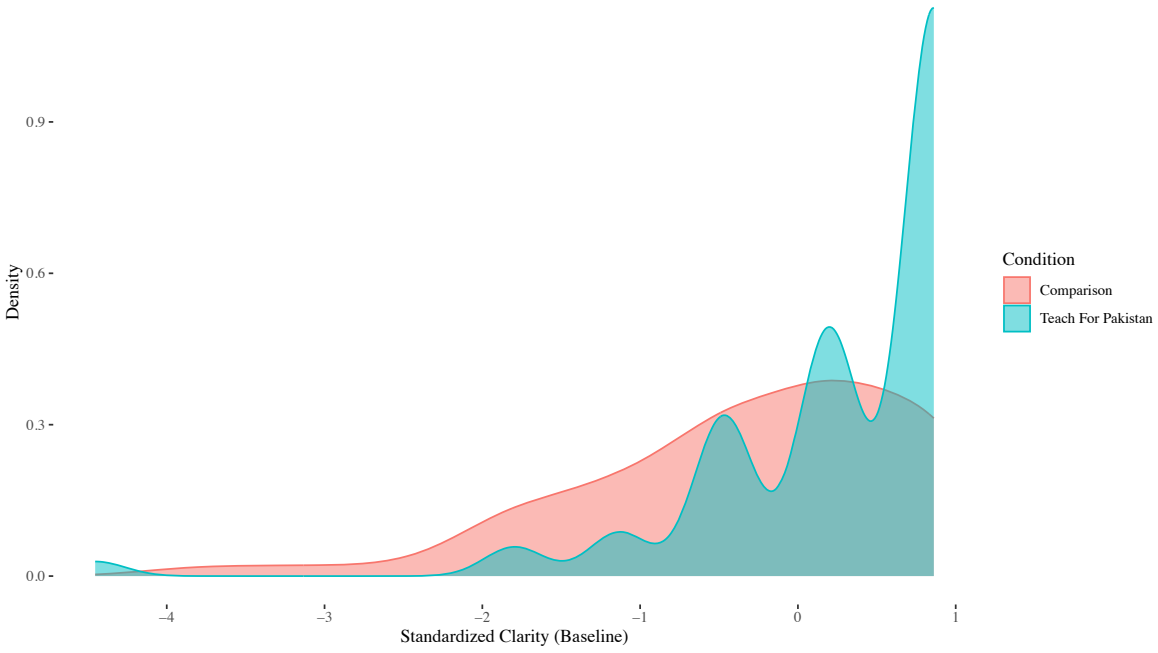


NOTE: We restricted our sample to teachers who participated in both baseline and follow-up data collections.

Figure A.13. Density Plot of Cognitive Activation

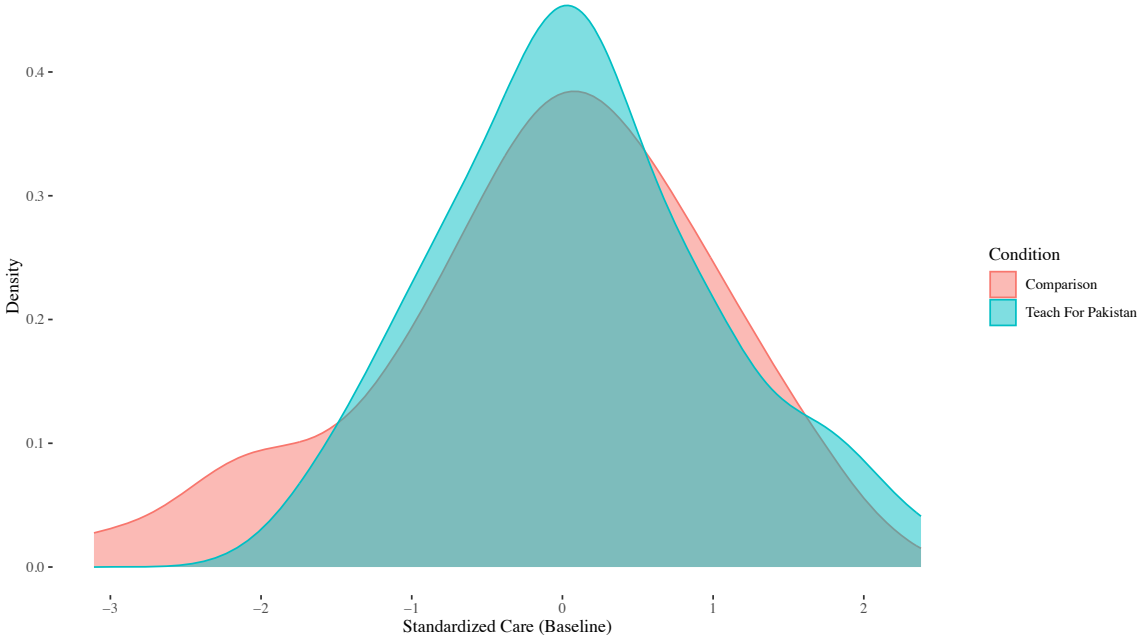
NOTE: We restricted our sample to teachers who participated in both baseline and follow-up data collections.

Figure A.14. Density Plot of Clarity



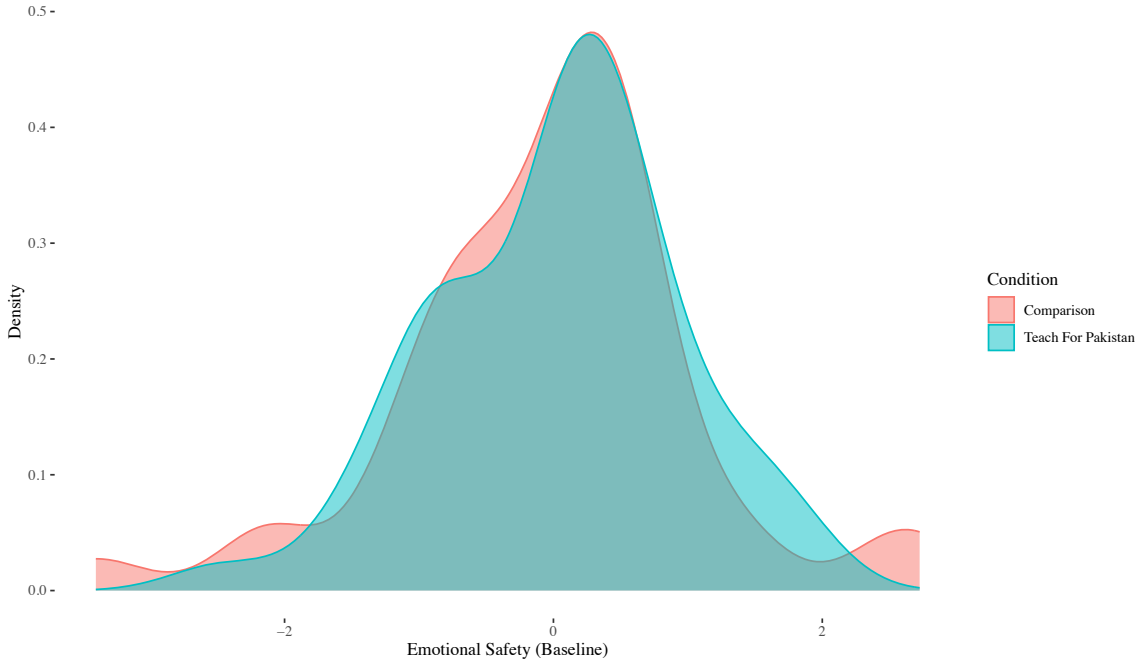
NOTE: We restricted our sample to teachers who participated in both baseline and follow-up data collections. We aggregated variables to the classroom level for analysis.

Figure A.15. Density Plot of Care



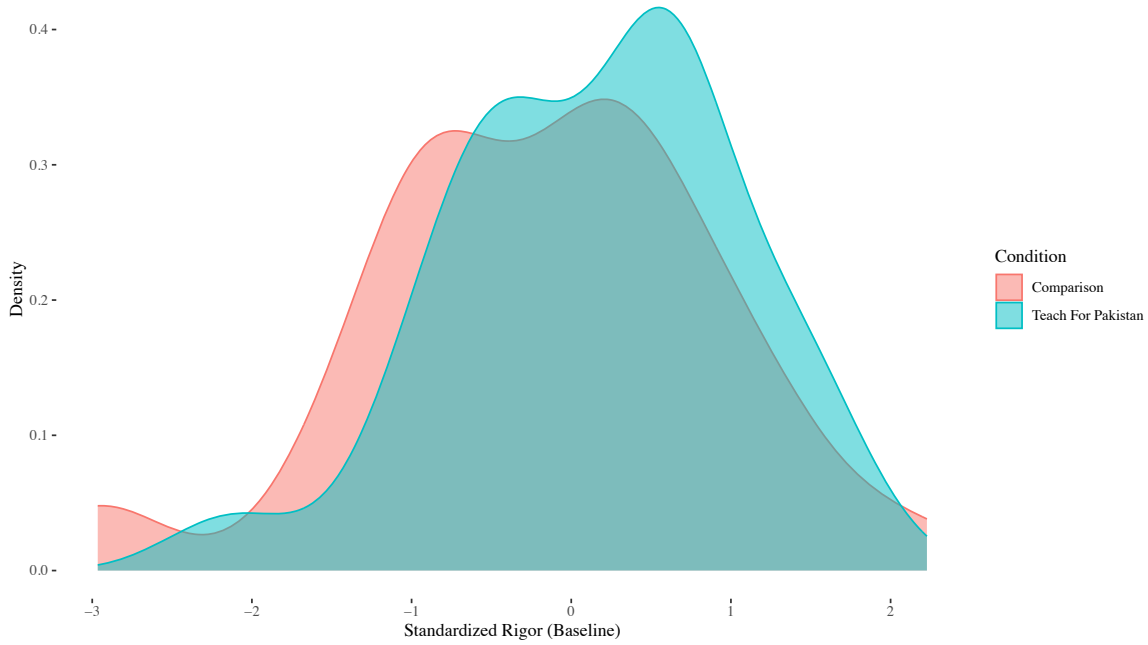
NOTE: We restricted our sample to teachers who participated in both baseline and follow-up data collections. We aggregated variables to the classroom level for analysis.

Figure A.16. Density Plot of Emotional Safety



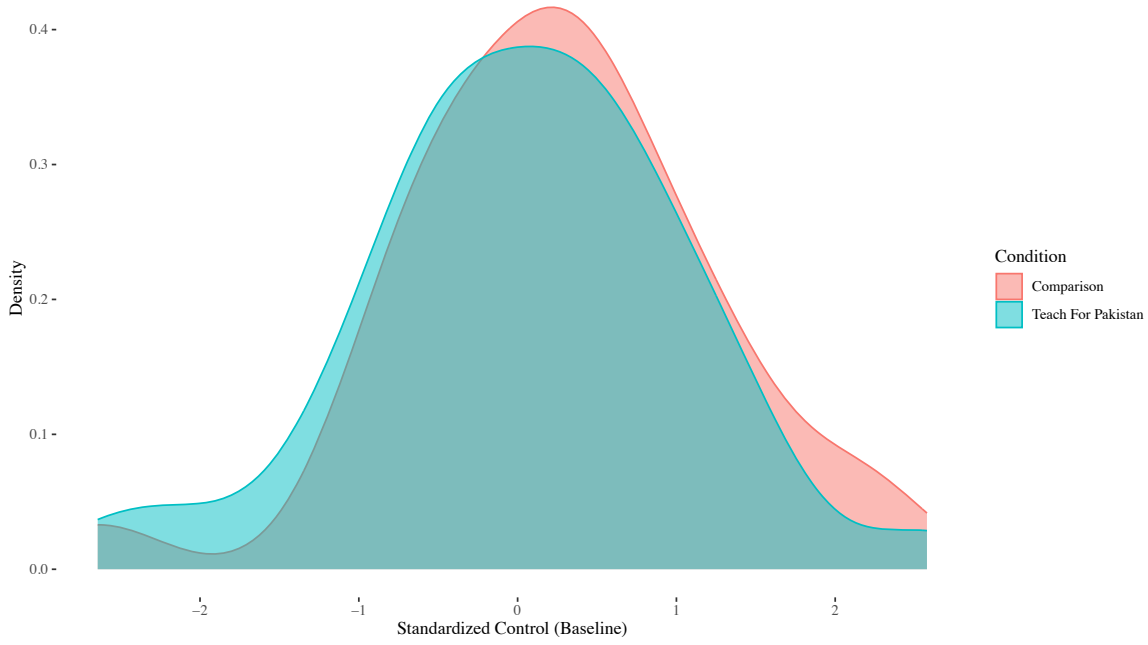
NOTE: We restricted our sample to teachers who participated in both baseline and follow-up data collections. We aggregated variables to the classroom level for analysis.

Figure A.17. Density Plot of Rigor



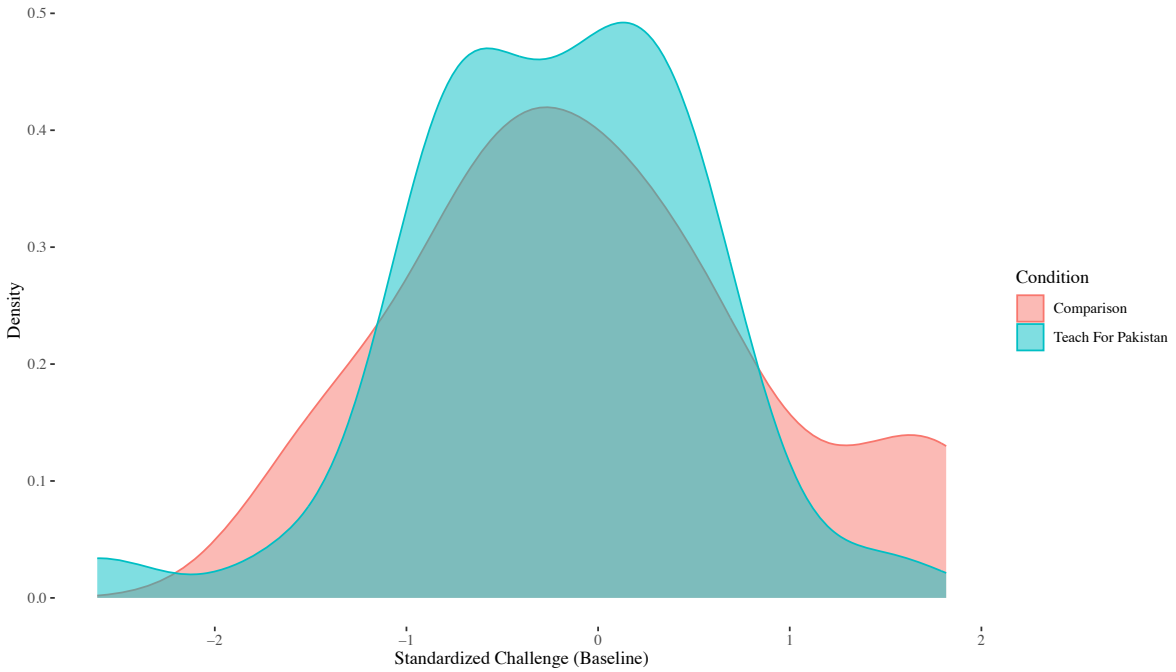
NOTE: We restricted our sample to teachers who participated in both baseline and follow-up data collections. We aggregated variables to the classroom level for analysis.

Figure A.18. Density Plot of Control



NOTE: We restricted our sample to teachers who participated in both baseline and follow-up data collections. We aggregated variables to the classroom level for analysis.

Figure A.19. Density Plot of Challenge



NOTE: We restricted our sample to teachers who participated in both baseline and follow-up data collections. We aggregated variables to the classroom level for analysis.

Standardized Mean Differences of the Unweighted Sample

In this section, we present the mean characteristics of the treatment and comparison classrooms and the SMD for each of the teacher and classroom characteristics included in our analytic models. As shown in Table A.2, many of the differences reported are greater than the 0.25 SMD that is used by the *What Works Clearinghouse Procedures and Standards Handbook* in quasi-experimental studies, and there are a few specific variables for which the differences between the TFP Fellows and the comparison teachers are especially pronounced. Because of the magnitude of these differences, we employed covariate balancing methods in our analyses of classroom-level outcomes.

In terms of background, treatment teachers are more likely to live in the communities in which they teach and to have substantially less teaching experience. In terms of survey-based variables describing teaching quality and the quality of the learning environment, treatment teachers have higher average ratings of their clarity, rigorous expectations, cognitive activation, self-efficacy, responsibility for learning, and collaboration than their comparison peers. Some of these differences are not surprising: For example, given the fact that the TFP program is designed to recruit early-career teachers, we would anticipate that the TFP Fellows in our sample would be less experienced than the teachers in comparison schools. Additionally, screening, selection, and training for TFP Fellows might be more selective than typical teacher induction processes, which could result in a sample of teachers who have stronger feelings of self-efficacy.

Table A.2. Baseline Equivalence of Teacher and Classroom Characteristics (Unweighted Analysis Sample)

Statistic	Treatment Mean (SD)	Comparison Mean (SD)	SMD
Teacher			
Male	0.28 (0.45)	0.33 (0.47)	−0.11
Live in community	0.18 (0.39)	0.36 (0.48)	−0.42
Worked as a teacher (years)	1.25 (1.14)	14.96 (7.98)	−2.41
Teaching quality and quality of the learning environment			
Control	−0.06 (1.03)	0.06 (0.97)	−0.12
Challenge	−0.10 (0.96)	0.13 (1.05)	−0.22
Rigorous expectations	0.15 (0.90)	−0.18 (1.10)	0.33
Emotional safety	0.04 (0.90)	−0.04 (1.12)	0.08
Care	0.11 (0.91)	−0.12 (1.09)	0.23
Clarity	0.30 (0.88)	−0.31 (1.03)	0.63
Cognitive activation	0.45 (0.96)	−0.46 (0.82)	1.01
Classroom management	0.07 (0.99)	−0.08 (1.01)	0.15
Assessment use	0.06 (1.03)	−0.06 (0.97)	0.12
Self-efficacy	0.27 (0.90)	−0.28 (1.03)	0.56
Responsibility for learning	0.48 (0.96)	−0.49 (0.78)	1.11
Collaboration	0.28 (0.95)	−0.29 (0.97)	0.59

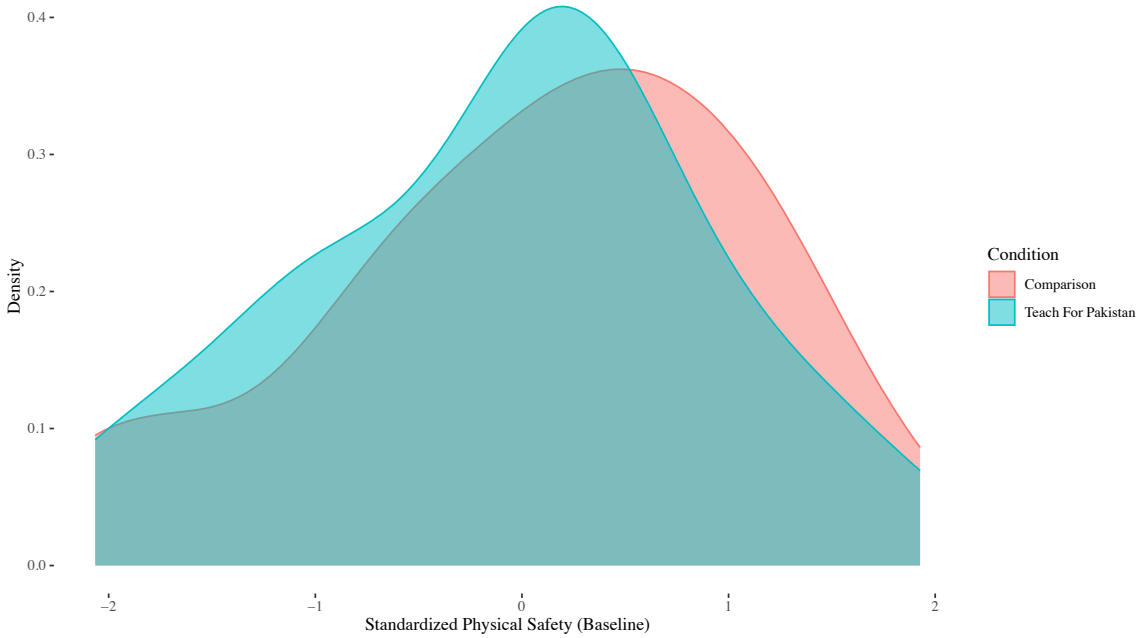
NOTE: We restricted our sample to teachers who participated in both baseline and follow-up data collections. Total $N = 146$. $N_{TFP} = 74$. $N_{COMPARISON} = 72$.

Baseline Equivalence of Schools

Common Support

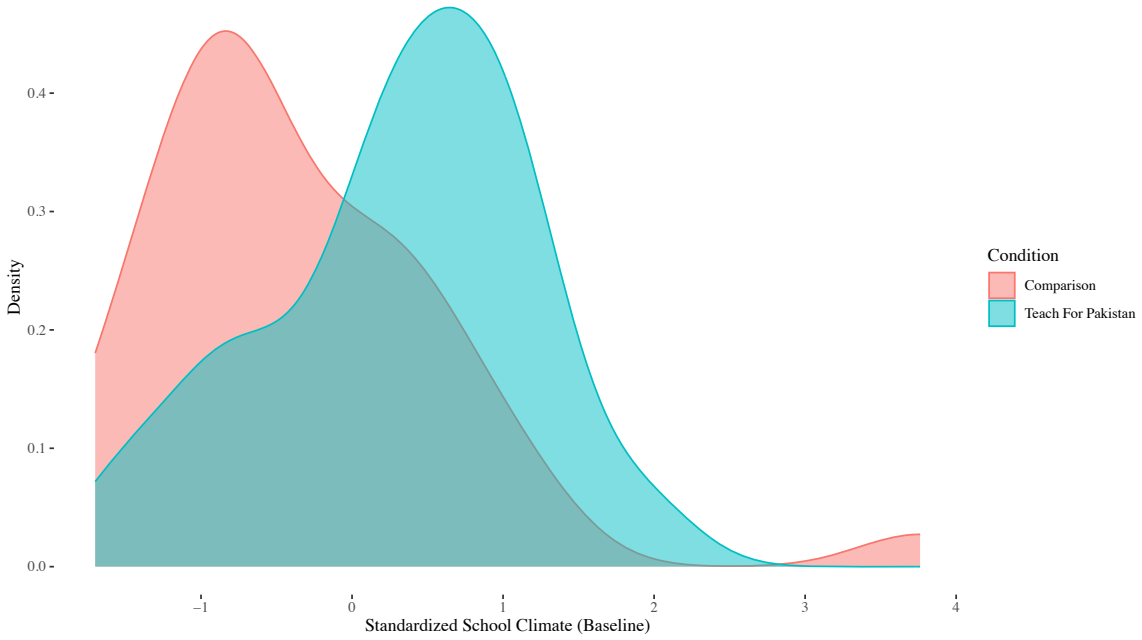
Figures A.20 through A.27 display density plots for the baseline measures of our outcome variables separately by quasi-experimental condition. The treatment schools are displayed in teal, and the comparison schools are displayed in pink. For each of these variables, there are noticeably larger differences in the distributions across each condition than there are for the student- or the classroom-level variables. However, there is still considerable overlap in the distribution, which provides evidence of common support.

Figure A.20. Density Plot of Physical Safety



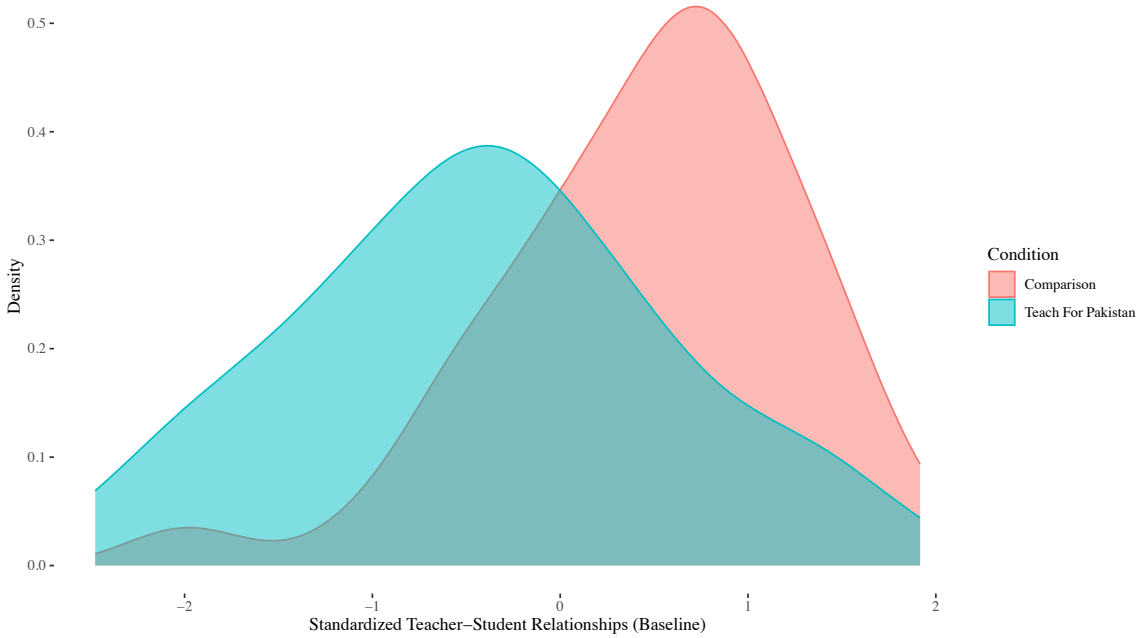
NOTE: We restricted our sample to schools that participated in both baseline and follow-up data collections. We aggregated variables to the school level for analysis.

Figure A.21. Density Plot of School Climate



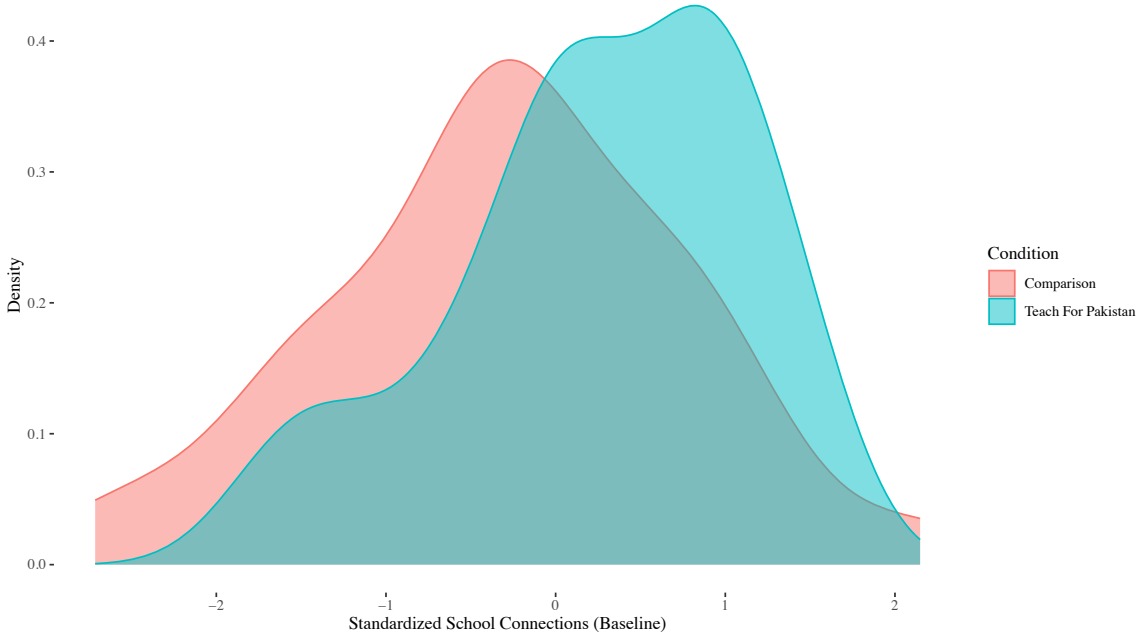
NOTE: We restricted our sample to schools that participated in both baseline and follow-up data collections. We aggregated variables to the school level for analysis.

Figure A.22. Density Plot of Teacher-Student Relationships



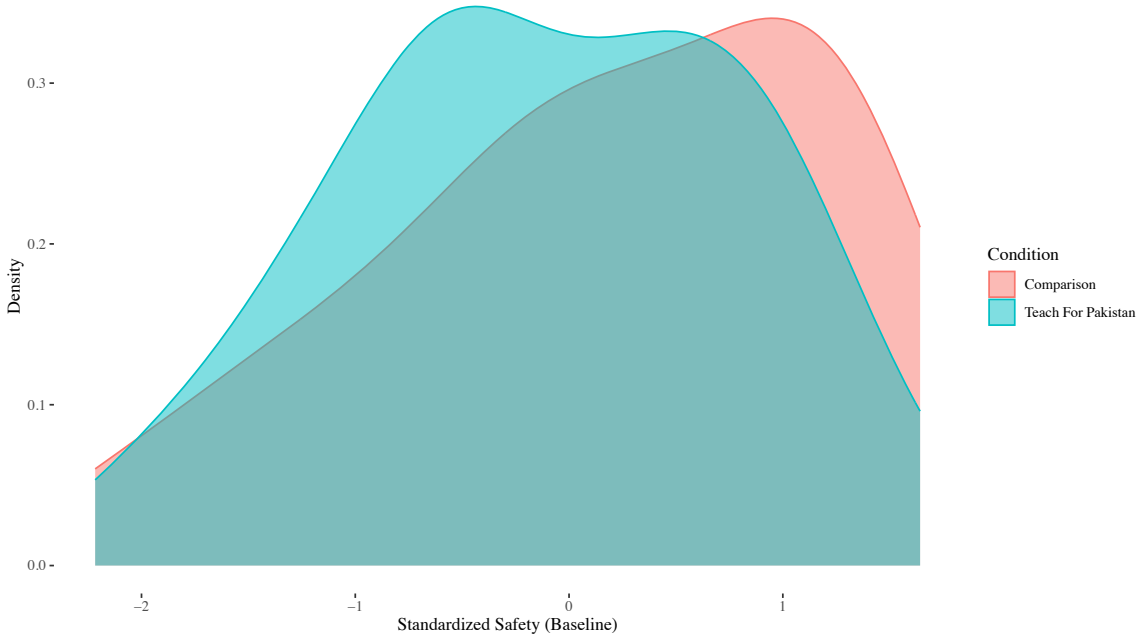
NOTE: We restricted our sample to schools that participated in both baseline and follow-up data collections. We aggregated variables to the school level for analysis.

Figure A.23. Density Plot of School Connections



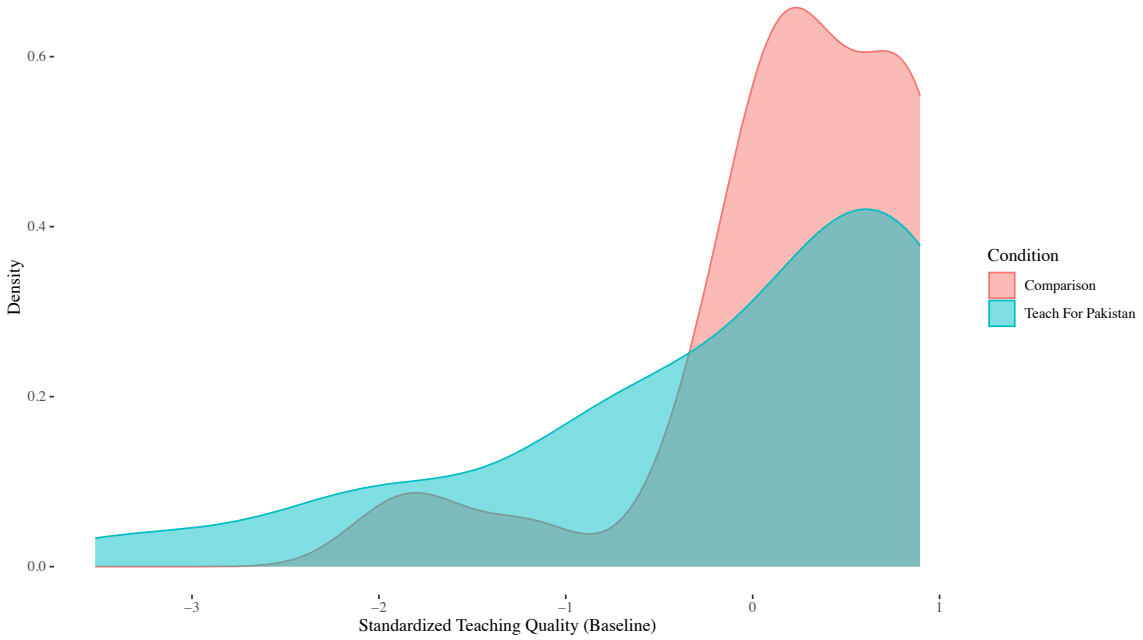
NOTE: We restricted our sample to schools that participated in both baseline and follow-up data collections. We aggregated variables to the school level for analysis.

Figure A.24. Density Plot of Safety



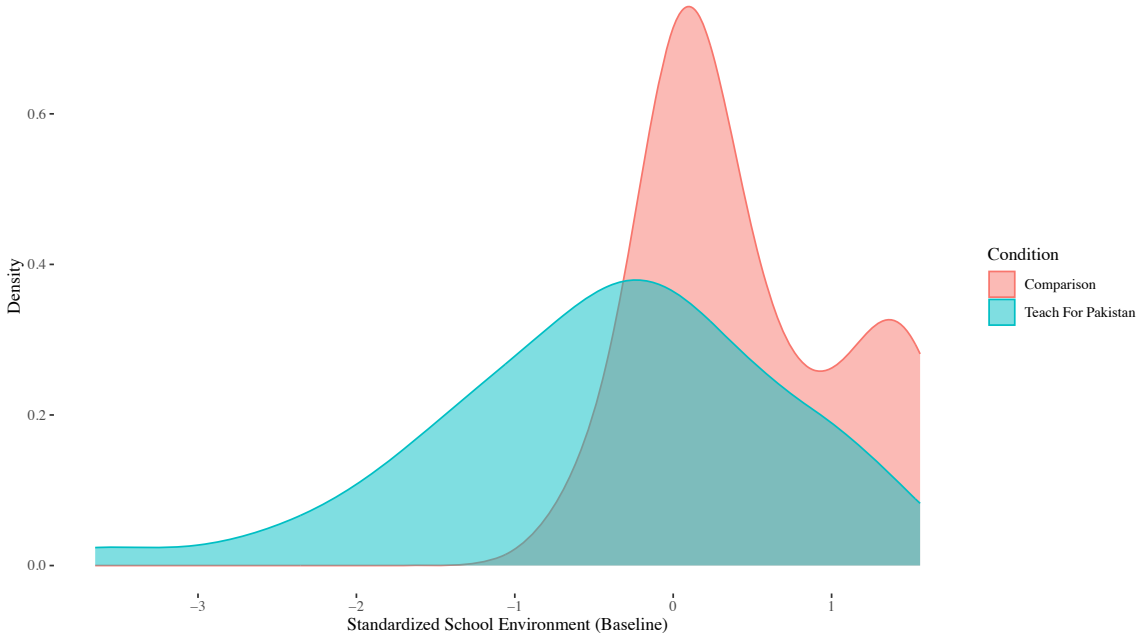
NOTE: We restricted our sample to schools that participated in both baseline and follow-up data collections. We aggregated variables to the school level for analysis.

Figure A.25. Density Plot of Teaching Quality



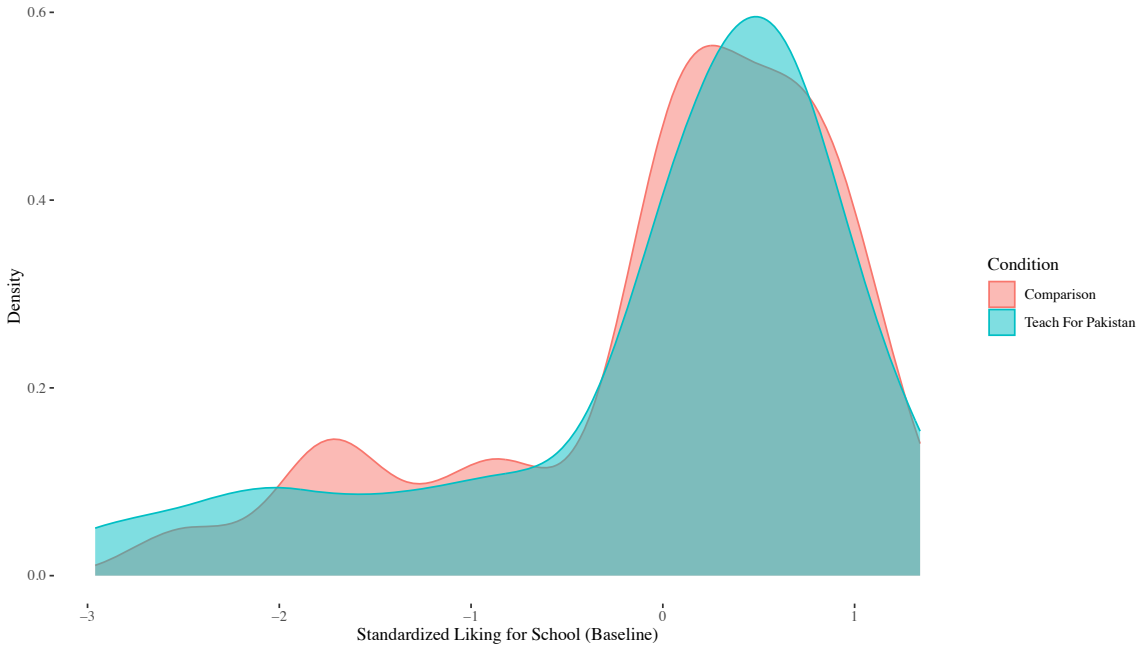
NOTE: We restricted our sample to schools that participated in both baseline and follow-up data collections. We aggregated variables to the school level for analysis.

Figure A.26. Density Plot of School Environment



NOTE: We restricted our sample to schools that participated in both baseline and follow-up data collections. We aggregated variables to the school level for analysis.

Figure A.27. Density Plot of Liking for School



NOTE: We restricted our sample to schools that participated in both baseline and follow-up data collections. We aggregated variables to the school level for analysis.

Standardized Mean Differences of the Unweighted Sample

In this section, we present the mean characteristics of the treatment (TFP) and comparison schools and report the SMD for each of the characteristics that are included in our analytic models. Nearly all of the differences are well above the 0.25 SMD that is used by the *What Works Clearinghouse Procedures and Standards Handbook* in quasi-experimental studies, as shown in Table A.3. Because of the magnitude of these differences, we employed covariate balancing methods in our analyses of classroom-level outcomes.

In particular, treatment schools are more likely to have higher ratings of school climate, school connections, and physical safety and lower ratings of safety (based on teacher survey reports), teacher-student relationships, and school environment.[26] Importantly, although we characterize these variables as school characteristics, they are based on a small sample of teachers and students in each school. Specifically, school connections, teacher-student relationships, school climate, physical safety, school environment, and teaching quality are all derived from the responses of the small sample of participating teachers in each school (in most cases, only two teachers), and safety and liking for school are derived from the responses of the students enrolled in a small sample of participating grades and classrooms.

Table A.3. Baseline Equivalence of School Characteristics (Unweighted Analysis Sample)

Statistic	Treatment Mean (SD)	Comparison Mean (SD)	SMD
School type			0.09
Boys	0.24	0.25	N/A
Coeducation	0.26	0.29	N/A
Girls	0.50	0.46	N/A
School community			
Liking for school	–0.25 (1.07)	0.03 (0.95)	–0.05
Safety	–0.10 (0.93)	0.17 (1.05)	–0.28
School connections	0.26 (0.87)	–0.29 (1.07)	0.56
Teacher-student relationships	–0.42 (1.00)	0.47 (0.80)	–0.98
School climate	0.31 (0.87)	–0.33 (1.04)	0.67
Physical safety	–0.07 (0.88)	–0.31 (1.03)	0.63
School environment	–0.42 (1.08)	0.48 (0.64)	–1.01
Teaching quality	–0.18 (1.18)	0.20 (0.72)	–0.38

NOTE: We restricted our sample to schools that participated in both baseline and follow-up data collections. Total $N = 80$. $N_{TFP} = 41$. $N_{COMPARISON} = 39$. N/A = not applicable.

[26] These ratings apply to constructs that are drawn from a variety of sources. Refer to Tables 3.1–3.3 in Chapter 3 for lists of sources and illustrative survey items related to each construct.

Appraising Balance After Weighting, Student-Level Analyses

As described previously, it is evident that prior to adjusting for propensity scores, the treatment and comparison students differed in terms of some background characteristics, particularly in terms of assessment scores in mathematics. Table A.4 shows the covariate balancing algorithms and included covariates for each of our analyses. Table A.5 shows the differences between treatment and comparison group respondents after applying the weights from the covariate balancing. Although the differences for some variables were larger than 0.25 standard deviations before weighting, the SMDs for all variables between the treatment and comparison groups were less than 0.25 standard deviations after weighting. Overall, Table A.5 shows that, after weighting, the treatment and comparison groups were equivalent on all baseline observed covariates.

Table A.4. Covariate Balancing Algorithms for Student Analyses

Statistic	Covariate Balancing Algorithm	Variables Included in Balancing
SEL skills		
Empathy	Logistic regression	Socioeconomic status, food insecurity, household size, repeated a grade, home language, gender, baseline empathy
Growth mindset	Logistic regression	Socioeconomic status, food insecurity, household size, repeated a grade, home language, gender, baseline growth mindset
Self-management	Logistic regression	Socioeconomic status, food insecurity, household size, repeated a grade, home language, gender, baseline self-management
Self-efficacy	Logistic regression	Socioeconomic status, food insecurity, household size, repeated a grade, home language, gender, baseline self-efficacy
Academic performance		
English	GBM	Socioeconomic status, food insecurity, household size, repeated a grade, home language, gender, baseline English
Mathematics	GBM	Socioeconomic status, food insecurity, household size, repeated a grade, home language, gender, baseline mathematics
Science	GBM	Socioeconomic status, food insecurity, household size, repeated a grade, home language, gender, baseline science

NOTE: The variables included in balancing are taken from the baseline survey.

Table A.5. Baseline Equivalence of Student Background Characteristics (Weighted Sample)

Statistic	Empathy (SMD)	Growth Mindset (SMD)	Self-Management (SMD)	Self-Efficacy (SMD)	English (SMD)	Mathematics (SMD)	Science (SMD)
SEL skills							
Empathy	0.00						
Growth mindset		0.02					
Self-management			0.00				
Self-efficacy				0.01			
Academic performance							
English					0.04		
Mathematics						0.01	
Science							0.04
Student characteristics							
Socioeconomic status	0.00	0.02	0.00	0.01	0.00	0.03	0.02
Food insecurity	0.00	0.01	0.00	0.00	0.04	0.03	0.06
Number of children in household	0.00	0.00	0.00	0.00	0.01	0.03	0.00
Repeated a grade	0.00	0.01	0.00	0.00	0.04	0.02	0.04
Gender (male)	0.01	0.02	0.00	0.01	0.04	0.02	0.06

NOTE: We restricted our sample to students who participated in both baseline and follow-up data collections.

Appraising Balance After Weighting, Classroom-Level Analyses

Similarly, for the classroom-level analyses, prior to adjusting for propensity scores, the treatment and comparison classrooms differed in terms of some baseline aspects, particularly teacher prior experience and aspects of teaching quality and the learning environment (e.g., baseline measures of the classroom-level outcomes). Table A.6 shows the covariate balancing algorithms and included covariates for each of our analyses. Tables A.7 and A.8 show the differences between treatment and comparison group respondents after applying the weights from the covariate balancing. Although the differences for some variables were considerably larger than 0.25 standard deviations before weighting, the SMDs between the treatment and comparison groups were generally less than 0.25 standard deviations after weighting. Importantly, the SMDs for all the baseline measures of the outcomes were less than 0.25 standard deviations. However, we were unable to achieve balance in terms of teacher background, and, in particular, there were persistent larger differences in teacher experience even after weighting. This is not surprising, given that TFP focuses explicitly on new teachers, and all TFP

Fellows in the study have either no experience or one year of experience. We were not able to place the same constraint on recruiting for the comparison group, so the teachers in that sample are considerably more experienced, on average. All attempts to achieve balance on experience resulted in weights that deteriorated the quality of the balance on other variables, including the baseline measures of the outcomes. For this reason, we note that selection bias based on experience might weaken our causal inferences about these variables, even though we included teacher experience as a covariate in all our regression models.

Table A.6. Covariate Balancing Algorithms for Classroom Analyses

Statistic	Covariate Balancing Algorithm	Variables Included in Balancing
Teacher self-reports		
Clarity	CBPS	School mean socioeconomic status, school mean food insecurity, school mean household size, school mean repeated a grade, percentage of Urdu speakers, school mean gender, teacher gender, teacher commute length, baseline clarity
Cognitive activation	EB	School mean socioeconomic status, school mean food insecurity, school mean household size, school mean repeated a grade, percentage of Urdu speakers, school mean gender, teacher gender, teacher commute length, baseline cognitive activation
Classroom management	CBPS	School mean socioeconomic status, school mean food insecurity, school mean household size, school mean repeated a grade, percentage of Urdu speakers, school mean gender, teacher gender, teacher commute length, baseline classroom management
Assessment use	CBPS	School mean socioeconomic status, school mean food insecurity, school mean household size, school mean repeated a grade, percentage of Urdu speakers, school mean gender, teacher gender, teacher commute length, baseline assessment use
Self-efficacy	CBPS	School mean socioeconomic status, school mean food insecurity, school mean household size, school mean repeated a grade, percentage of Urdu speakers, school mean gender, teacher gender, teacher commute length, baseline self-efficacy
Responsibility for learning	EB	School mean socioeconomic status, school mean food insecurity, school mean household size, school mean repeated a grade, percentage of Urdu speakers, school mean gender, teacher gender, teacher commute length, baseline responsibility for learning

Statistic	Covariate Balancing Algorithm	Variables Included in Balancing
Collaboration	CBPS	School mean socioeconomic status, school mean food insecurity, school mean household size, school mean repeated a grade, percentage of Urdu speakers, school mean gender, teacher gender, teacher commute length, baseline collaboration
Student reports		
Control	GBM	School mean socioeconomic status, school mean food insecurity, school mean household size, school mean repeated a grade, percentage of Urdu speakers, school mean gender, teacher gender, teacher commute length, baseline control
Challenge	CBPS	School mean socioeconomic status, school mean food insecurity, school mean household size, school mean repeated a grade, percentage of Urdu speakers, school mean gender, teacher gender, teacher commute length, baseline challenge
Rigorous expectations	CBPS	School mean socioeconomic status, school mean food insecurity, school mean household size, school mean repeated a grade, percentage of Urdu speakers, school mean gender, teacher gender, teacher commute length, baseline rigorous expectations
Emotional safety	CBPS	School mean socioeconomic status, school mean food insecurity, school mean household size, school mean repeated a grade, percentage of Urdu speakers, school mean gender, teacher gender, teacher commute length, baseline emotional safety
Care	CBPS	School mean socioeconomic status, school mean food insecurity, school mean household size, school mean repeated a grade, percentage of Urdu speakers, school mean gender, teacher gender, teacher commute length, baseline care

NOTE: The variables included in balancing are taken from the baseline surveys.

Table A.7. Baseline Equivalence for Classroom Analyses Based on Student Reports (Weighted Sample)

Statistic	Control (SMD)	Challenge (SMD)	Rigorous Expectations (SMD)	Emotional Safety (SMD)	Care (SMD)
School					
Socioeconomic status	0.04	0.16	0.02	0.08	0.03
Food insecurity	0.06	0.32	0.12	0.04	0.23
Household size	0.26	0.17	0.15	0.07	0.15
Repeated a grade	0.03	0.01	0.02	0.21	0.03
Urdu	0.04	0.02	0.05	0.03	0.01
Gender	0.01	0.08	0.05	0.01	0.04
Teacher					
Male	0.00	0.02	0.05	0.02	0.05
Live in community	0.74	0.70	0.84	0.75	0.56
Worked as a teacher (years)	2.51	2.72	2.46	2.50	2.72
Teaching quality and quality of the learning environment					
Control	0.01				
Challenge		0.17			
Rigorous expectations			0.08		
Emotional safety				0.02	
Care					0.17

NOTE: We restricted our sample to teachers who participated in both baseline and follow-up data collections.

Table A.8. Baseline Equivalence for Classroom Analyses Based on Teacher Reports (Weighted Sample)

Statistic	Clarity (SMD)	Cognitive Activation (SMD)	Classroom Management (SMD)	Assessment Use (SMD)	Self-Efficacy (SMD)	Responsibility for Learning (SMD)	Collaboration (SMD)
School							
SES	0.02	0.00	0.02	0.02	0.05	0.00	0.05
Food insecurity	0.14	0.00	0.12	0.08	0.16	0.00	0.06
Household size	0.02	0.00	0.08	0.11	0.05	0.00	0.09
Repeated a grade	0.16	0.00	0.03	0.04	0.05	0.00	0.05
Male	0.01	0.00	0.04	0.02	0.04	0.00	0.04
Urdu	0.04	0.00	0.02	0.03	0.02	0.00	0.02
Teacher							
Male	0.02	0.00	0.05	0.04	0.03	0.00	0.00
Live in community	0.03	0.00	0.04	0.04	0.04	0.00	0.00
Worked as a teacher	2.57	2.72	2.71	2.64	2.78	2.84	2.71
Teaching quality and quality of the learning environment							
Clarity	0.05						
Cognitive activation		0.00					
Classroom management			0.00				
Assessment use				0.06			
Self-efficacy					0.06		
Responsibility for Learning						0.23	
Collaboration							0.19

NOTE: We restricted our sample to schools participating in both baseline and follow-up data collections.

Appraising Balance After Weighting, School-Level Analyses

Prior to adjusting for propensity scores, the treatment and comparison classrooms differed in terms of some baseline measures of the outcome variables. Table A.9 shows the covariate balancing algorithms and included covariates for each of our analyses. Note that a much smaller set of covariates was included in the balancing. This is because the school-level sample was relatively small, so we reduced the number of covariates in the balancing algorithms accordingly. Tables A.10 and A.11 show the differences between treatment and comparison groups after applying the weights from the covariate balancing. Although the differences for some variables were considerably larger than 0.25

standard deviations before weighting, the SMDs between the treatment and comparison groups were generally less than 0.25 standard deviations after weighting. The differences for two outcomes (teacher-student relationships and the school environment) were still considerably larger than 0.25. For this reason, and because our balancing was based on a smaller set of covariates, we note that selection bias might weaken our causal inferences about these outcomes.

Table A.9. Covariate Balancing Algorithms for School Analyses

Statistic	Covariate Balancing Algorithm	Covariates Included in Balancing
Principal self-reports		
School environment	CBPS	School mean socioeconomic status, school mean food insecurity, percentage of Urdu speakers, school mean gender, baseline school environment
Teaching quality	CBPS	School mean socioeconomic status, school mean food insecurity, percentage of Urdu speakers, school mean gender, baseline teaching quality
Teacher reports		
Teacher-student relationships	CBPS	School mean socioeconomic status, school mean food insecurity, percentage of Urdu speakers, school mean gender, baseline teacher-student relationships
School climate	CBPS	School mean socioeconomic status, school mean food insecurity, percentage of Urdu speakers, school mean gender, baseline school climate
Physical safety	CBPS	School mean socioeconomic status, school mean food insecurity, percentage of Urdu speakers, school mean gender, baseline physical safety
Student reports		
Liking for school	CBPS	School mean socioeconomic status, school mean food insecurity, percentage of Urdu speakers, school mean gender, baseline liking for school
Safety	CBPS	School mean socioeconomic status, school mean food insecurity, percentage of Urdu speakers, school mean gender, baseline safety
School connections	CBPS	School mean socioeconomic status, school mean food insecurity, percentage of Urdu speakers, school mean gender, baseline school connections

NOTE: The variables included in balancing are taken from the baseline survey.

Table A.10. Baseline Equivalence for School Analyses Based on Student Reports (Weighted Sample)

Statistic	Liking for School (SMD)	Safety (SMD)	School Connections (SMD)
Liking for school	0.07		
Safety		0.06	
School connections			0.21

NOTE: We restricted our sample to teachers participating in both baseline and follow-up data collections.

Table A.11. Baseline Equivalence for Classroom Analyses Based on Teacher and Principal Reports (Weighted Sample)

Statistic	Teacher-Student Relationships (SMD)	School Climate (SMD)	Physical Safety (SMD)	School Environment (SMD)	Teaching Quality (SMD)
Teacher-student relationships	0.46				
School climate		0.21			
Physical safety			0.00		
School environment				0.67	
Teaching quality					0.19

NOTE: We restricted our sample to teachers and schools participating in both baseline and follow-up data collections.

Appendix B

Data Analysis Methods

Quantitative Study Student-Level Analysis

Estimates of the impact of TFP were obtained by comparing the outcomes of treatment- and comparison-group students while controlling for the small differences at baseline between the two groups. We used a canonical difference-in-differences model with two periods and a single treatment (Roth et al., 2023). The treatment effect was estimated using an ordinary least squares model that sandwich-estimated standard errors to account for the clustering of students within schools (White, 1980):

$$y_{ics} = \beta_0 + \beta_1 T_{cs} + X_{ics}'\lambda + Z_{ics}'\delta + \gamma + e_{ics}.$$

In this equation,

- y_{ics} is the outcome for student *i* in classroom *c* in school *s*.
- T_{cs} is a treatment indicator (1 = enrolled in a TFP school, and 0 = otherwise).
- X_{ics} is a vector of baseline student covariates, including socioeconomic status, food insecurity, number of children in the household, home language, and gender.
- Z_{ics} is a vector of baseline measures of SEL skills or academic proficiency, including self-efficacy, growth mindset, social awareness, mathematics, English, and science.
- e_{ics} is a residual term with mean 0 and variance σ^2.
- γ is a vector of school block (matched pair) fixed effects.[27]

We interpret $\hat{\beta}_1$, the estimate of β_1, as an estimate of TFP impact.

Quantitative Study Classroom-Level Analysis

Some of our classroom-level outcome variables (care, challenge, control, rigor, and emotional safety) were derived from scales on the student survey. Therefore, the first step in each of these analyses was to aggregate the student survey responses to create classroom-level variables. For the teacher self-report variables (clarity, cognitive activation, classroom management, assessment use, self-efficacy, responsibility for learning, and collaboration), this step was not necessary. All regressions were conducted using classroom-level variables.

[27] Our models do not incorporate grade-level or school gender fixed effects because these variables are collinear with school block (matched pair) fixed effects. The student-level analyses drop students in two treatment schools that do not have a matched comparison school.

The treatment effect was estimated using an ordinary least squares model, weighted by the estimated balancing weight, and using sandwich-estimated standard errors to account for the clustering of classrooms within schools:

$$y_{cs} = \beta_0 + \beta_1 T_{cs} + X_{cs}'\lambda + W_{cs}'\eta + \beta_2 Z_{cs} + \gamma + e_{cs}.$$

In this equation,

- y_{cs} is the outcome for classroom *c* in school *s*.
- T_{cs} is a treatment indicator (1 = enrolled in a TFP school, and 0 = otherwise).
- X_{cs} is a vector of baseline school-level covariates, including school-average socioeconomic status, percentage experiencing food insecurity, household size, percentage male, percentage of students who repeated a grade, and class size.
- W_{cs} is a vector of baseline teacher covariates, including gender, teaching experience, and whether the teacher lives in the community.
- Z_{cs} is a baseline measure of the outcome.
- e_{cs} is a residual term with mean 0 and variance σ^2.
- γ is a vector of school block (matched pair) fixed effects.

We interpret $\hat{\beta}_1$, the estimate of β_1, as an estimate of TFP impact.

Quantitative Study School-Level Analysis

For the school-level analyses, three outcome variables were measured using the teacher survey (teacher-student relationships, physical safety, and school climate), three outcomes were measured using the student survey (liking for school, safety, and school connections), and two outcome variables were measured using the principal survey (school environment and teaching quality). Therefore, the first step in the school-level analysis was to aggregate the student survey responses to create school-level variables, and, analogously, to aggregate the teacher survey responses to create school-level variables. All regressions were conducted using school-level variables.

The treatment effect was estimated using an ordinary least squares model, weighted by the estimated balancing weight:

$$y_s = \beta_0 + \beta_1 T_s + X_s'\lambda + W_s'\eta + \beta_2 Z_s + e_s.$$

In this equation,

- y_s is the outcome for school *s*.
- T_s is a treatment indicator (1 = enrolled in a TFP school, and 0 = otherwise).
- X_s is a vector of baseline school covariates, including school-average socioeconomic status, food insecurity, gender, and percentage of students who repeated a grade.
- Z_s is a baseline measure of the outcome.
- e_s is a residual term with mean 0 and variance σ^2.

We interpret $\hat{\beta}_1$, the estimate of β_1, as an estimate of TFP impact.

Qualitative Study Analysis

Table B.1 shows the mapping of evaluation questions to interview protocol questions to the coding scheme for the qualitative study and how they relate to the TFP theory of change.

Table B.1. Qualitative Study: Mapping of Evaluation Questions, Interview Questions, Teach For Pakistan Theory of Change, and Coding Scheme

Evaluation Question	Sub-Questions	Key Interview and Focus Group Protocol Questions	Sample Second-Level Codes in Coding Scheme	Relevant TFP Theory of Change Constructs
What do various stakeholders (e.g., principals, other teachers, parents, students, and TFP Fellows themselves) perceive as the contributions of TFP Fellows on whole-child development, including student academic learning and SEL outcomes?	How do various stakeholders perceive the contributions of TFP Fellows on students' **academic learning?**	What would you say are the main contributions of the TFP Fellows to students' academic and cognitive development?	Academic development • Reading skills • Writing • Numeracy skills • STEM-related skills • Critical thinking skills	Outcomes: mastery of key subject areas and academic skills (i.e., numeracy, literacy)
	How do various stakeholders perceive the contributions of TFP Fellows on students' **social-emotional outcomes?**	What would you say are the main contributions of the TFP Fellows to students' **intrapersonal** (individual) social-emotional development that is unique to TFP Fellows? What would you say are the main contributions of the TFP Fellows to students' **interpersonal** development?	Social and emotional skills • Communication skills • Empathy and compassion • Collective responsibility • Self-awareness • Agency and autonomy • Self-confidence • Grit and resilience	Outcomes: increased agency, grit, and sense of collective responsibility
How do stakeholders characterize the teaching approach and classroom environment of TFP Fellows?	How do various stakeholders characterize the **teaching approach** of TFP Fellows?	How have TFP Fellows implemented their vision of rigorous classroom instruction? What do you perceive as the instructional strengths of the TFP Fellows? How have TFP Fellows engaged students in their learning?	Instructional approach • Mastery based • Rigorous academic expectations • Supportive • Differentiated • Active, creative pedagogies • Student-centered, empowering	Outputs: establish an ambitious classroom vision and enact it through the design and delivery of rigorous lesson plans, tailor the curriculum to start teaching from students' actual learning levels, and track students' data and progress and adapt lesson and learning plans

Evaluation Question	Sub-Questions	Key Interview and Focus Group Protocol Questions	Sample Second-Level Codes in Coding Scheme	Relevant TFP Theory of Change Constructs
	How do various stakeholders characterize the **classroom environment** created by TFP Fellows?	How do TFP Fellows cultivate grit in students? To what extent do students demonstrate collective responsibility by encouraging and supporting each other? What do the TFP Fellows do to foster this?	Classroom environment • Peer support and collaboration • Sense of community • Student participation and engagement in learning • Opportunities for student voice, autonomy, and leadership • Conflict prevention and resolution	Outputs: believe all students can succeed and have resilience, engage students with understanding and compassion, center the learning environment on students' well-being and development
What do stakeholders perceive as the contributions of TFP Fellows to the school community, including working with stakeholders to support students?	What is the quality of TFP Fellows' **relationships with different stakeholders** (e.g., the principal, other teachers, students, and parents)?	How have TFP Fellows engaged students' parents in their children's learning? How have TFP Fellows worked together with you and other stakeholders to understand and support students?	Relationship description • Friendly, open • Supportive, helpful • Trusting, respectful • Neutral • Strained, difficult	Outputs: increase the engagement of parents in students' success and in the school overall, work closely with parents and school staff to deepen a contextual understanding of students and to build an ecosystem of support for students
	How do various stakeholders perceive the **contributions of TFP Fellows on the school community**?	What contributions have TFP Fellows made to the school community? How have TFP Fellows worked with you and other stakeholders to improve school climate?	Main contributions • Materials and resources • Pedagogical approaches • Foster teacher collaboration • Increase parent engagement • School-wide initiatives • Improve school climate	Outputs: work with other stakeholders and leaders to build partnerships and improve school climate

NOTE: STEM = science, technology, engineering, and mathematics.

Abbreviations

CBPS	Covariate Balancing Propensity Score
CoBWeb	Covariate Balancing and Weighting Web App
CORE	California Office to Reform Education
CPP	community partnership project
EB	entropy balancing
FDE	Federal Directorate of Education
GBM	general boosted model
OECD	Organisation for Economic Co-operation and Development
RCons	Research Consultants
SEL	social and emotional learning
SMD	standardized mean difference
TALIS	Teaching and Learning International Survey
TFP	Teach For Pakistan
UNICEF	United Nations International Children's Emergency Fund

References

ASER Pakistan, homepage, undated. As of December 18, 2024: https://aserpakistan.org

Austin, G., and Duerr, M., *Guidebook for the California Healthy Kids Survey*: Part III, *School Climate Survey for Teachers and Other Staff*, WestEd, 2005.

Bernard, H. Russell, Amber Wutich, and Gery W. Ryan, *Analyzing Qualitative Data: Systematic Approaches*, 2nd ed., Sage Publications, 2016.

Bryk, Anthony S., Penny Bender Sebring, Elaine Allensworth, Stuart Luppescu, and John Q. Easton, *Organizing Schools for Improvement: Lessons from Chicago*, University of Chicago Press, 2010.

Burtless, Gary, "The Case for Randomized Field Trials in Economic and Policy Research," *Journal of Economic Perspectives*, Vol. 9, No. 2, Spring 1995.

Chen, Gilad, Stanley M. Gully, and Dov Eden, "Validation of a New General Self-Efficacy Scale," *Organizational Research Methods*, Vol. 4, No. 1, January 2001.

Chicago Consortium on School Research, *5Essentials School Report: The 5Essentials Full Report*, University of Chicago, 2011.

Cipriano, Christina, Michael J. Strambler, Lauren H. Naples, Cheyeon Ha, Megan Kirk, Miranda Wood, Kaveri Sehgal, Almut K. Zieher, Abigail Eveleigh, Michael McCarthy, et al., "The State of Evidence for Social and Emotional Learning: A Contemporary Meta-Analysis of Universal School-Based SEL Interventions," *Child Development*, Vol. 94, No. 5, 2023.

Clark, Melissa A., Eric Isenberg, Albert Y. Liu, Libby Makowsky, and Marykate Zukiewicz, *Impacts of the Teach For America Investing in Innovation Scale-Up*, Mathematica Policy Research, Inc., 2017.

Consulate General of Pakistan, Los Angeles, "About Pakistan," webpage, undated. As of December 18, 2024: https://pakconsulatela.org/about-pakistan

Creswell, John W., and J. David Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, Sage Publications, 2017.

Denzin, Norman K., ed., *Sociological Methods: A Sourcebook*, 1st ed., Routledge, 2006.

Digital Promise, *Learning in the 21st Century: How the American Public, Parents, and Teachers View Educational Priorities and How to Achieve Them*, 2021.

Duckworth, Angela L., and David Scott Yeager, "Measurement Matters: Assessing Personal Qualities Other Than Cognitive Ability for Educational Purposes," *Educational Researcher*, Vol. 44, No. 4, 2015.

EdInstruments, "SELweb," webpage, Annenberg Institute, undated. As of February 3, 2025: https://edinstruments.org/about

Ferguson, Ronald F., *Student Perceptions of Teaching Effectiveness*, discussion brief, National Center for Teacher Effectiveness and the Achievement Gap Initiative, Harvard University, October 14, 2010.

Fricke, Hans, Susanna Loeb, Robert H. Meyer, Andrew B. Rice, Libby Pier, and Heather Hough, "Stability of School Contributions to Student Social-Emotional Learning Gains," *American Journal of Education*, Vol. 128, No. 1, November 2021.

Grant, Sean, Laura S. Hamilton, Stephani L. Wrabel, Celia J. Gomez, Anamarie Whitaker, Jennifer T. Leschitz, Fatih Unlu, Emilio R. Chavez-Herrerias, Garrett Baker, Mark Barrett, Mark Harris, and Alyssa Ramos, *Social and Emotional Learning Interventions Under the Every Student Succeeds Act: Evidence Review*, RAND Corporation, RR-2133-WF, 2017. As of December 18, 2024: https://www.rand.org/pubs/research_reports/RR2133.html

Guest, Greg, Kathleen M. MacQueen, and Emily E. Namey, *Applied Thematic Analysis*, Sage Publications, 2011.

Hart, Shelley R., Kaitlyn Stewart, and Shane R. Jimerson, "The Student Engagement in Schools Questionnaire (SESQ) and the Teacher Engagement Report Form-New (TERF-N): Examining the Preliminary Evidence," *Contemporary School Psychology*, Vol. 15, 2011.

Imai, Kosuke, and Marc Ratkovic, "Covariate Balancing Propensity Score," *Journal of the Royal Statistical Society*, Series B: *Statistical Methodology*, Vol. 76, No. 1, January 2014.

International Trade Administration, "Pakistan—Country Commercial Guide: Education," webpage, U.S. Department of Commerce, January 12, 2024. As of December 19, 2024: https://www.trade.gov/country-commercial-guides/pakistan-education

Kershner, K., E. Schmidt, and R. Angelova, *Teach for Bulgaria Final Evaluation Report: 2018–2019*, Research and Evaluation International, 2019.

Kraft, Matthew A., "Interpreting Effect Sizes of Education Interventions," *Educational Researcher*, Vol. 49, No. 4, May 2020.

Kunter, Mareike, Uta Klusmann, Jürgen Baumert, Dirk Richter, Thamar Voss, and Axinja Hachfeld, "Professional Competence of Teachers: Effects on Instructional Quality and Student Development," *Journal of Educational Psychology*, Vol. 105, No. 3, 2013.

Lavado, Pablo, and Renzo Guzmán, *Impact Evaluation of Enseña Perú*, Universidad del Pacífico, October 2020.

Lincoln, Yvonna S., and Norman K. Denzin, eds., *Turning Points in Qualitative Research: Tying Knots in a Handkerchief*, AltaMira Press, 2003.

Lincoln, Yvonna S., and Egon G. Guba, *Naturalistic Inquiry*, Sage Publications, 1985.

Markoulidakis, Andreas, Peter Holmans, Philip Pallmann, Monica Busse, and Beth-Ann Griffin, "CoBWeb: A User-Friendly Web Application to Estimate Causal Treatment Effects from Observational Data Using Multiple Algorithms," *arXiv*, 2021.

McLean, Dawson, and Jack Worth, *The Impact of the Teach First Training Programme on Schools and Pupils*, National Foundation for Education Research, 2023.

Mihaly, Kata, Jonathan Schweig, Elaine Lin Wang, and Sabrina Lee, *Teach For Nigeria Evaluation: Quantitative and Qualitative Study Findings*, RAND Corporation, RR-A1870-1, 2024. As of December 18, 2024: https://www.rand.org/pubs/research_reports/RRA1870-1.html

Miles, Matthew B., A. Michael Huberman, and Johnny Saldaña, *Qualitative Data Analysis: A Methods Sourcebook*, 3rd ed., Sage Publications, 2014.

Ministry of Federal Education and Professional Training, *National Curriculum of Pakistan 2022*, Government of Pakistan, 2022.

Moore, Rhiannon, *The Design of the 2016–17 Young Lives School Survey in India*, Young Lives, Technical Note 37, December 2016.

National Center for Education Evaluation at the Institute of Education Sciences, *What Works Clearinghouse Procedures and Standards Handbook*, Version 5.0, U.S. Department of Education, 2022.

National Center for Education Statistics, "ED School Climate Surveys (EDSCLS)," webpage, undated-a. As of December 18, 2024:
https://nces.ed.gov/surveys/edscls/index.asp

National Center for Education Statistics, "Schools and Staffing Survey (SASS)," webpage, undated-b. As of December 18, 2024:
https://nces.ed.gov/surveys/sass

Nye, Barbara, Spyros Konstantopoulos, and Larry V. Hedges, "How Large Are Teacher Effects?" *Educational Evaluation and Policy Analysis*, Vol. 26, No. 3, Fall 2004.

OECD—*See* Organisation for Economic Co-operation and Development.

Organisation for Economic Co-operation and Development, *TALIS 2008 Technical Report*, 2008.

Organisation for Economic Co-operation and Development, *TALIS 2013 Technical Report*, 2013.

Organisation for Economic Co-operation and Development, *TALIS 2018 Technical Report*, 2018.

Palinkas, Lawrence A., Sarah M. Horwitz, Carla A. Green, Jennifer P. Wisdom, Naihua Duan, and Kimberly Hoagwood, "Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research," *Administration and Policy in Mental Health and Mental Health Services Research*, Vol. 42, No. 5, September 2015.

Panorama Education, *Validity Brief: Panorama Student Survey*, 2015.

Peña, Pablo A., and Armando Chacón, "Fostering Student Socioemotional Skills in Mexico," Microanalitica, 2017.

Roth, Jonathan, Pedro H. C. Sant'Anna, Alyssa Bilinski, and John Poe, "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature," *Journal of Econometrics*, Vol. 235, No. 2, August 2023.

Ryan, Gery W., and H. Russell Bernard, "Techniques to Identify Themes," *Field Methods*, Vol. 15, No. 1, February 2003.

Spier, Elizabeth, *Alaska School Climate and Connectedness Survey: 2016 Statewide Report*, American Institutes for Research, November 2016.

Teach For All, "Our Purpose," webpage, undated. As of December 18, 2024:
https://teachforall.org/our-purpose

Teach For All, *Global Teacher and Teacher Coach Study*, March 2023.

Teach For Pakistan, homepage, undated-a. As of December 19, 2024:
https://iteachforpakistan.org

Teach For Pakistan, "Fellowship FAQs," webpage, undated-b. As of January 9, 2025: https://iteachforpakistan.org/fellowship-faqs/

Teach For Pakistan, "Impact on Students," webpage, undated-c. As of February 3, 2025: https://iteachforpakistan.org/on-students/

Teach For Pakistan, "Teach For Pakistan Fellowship," internal document shared with the authors, undated-d.

Teach For Pakistan, "Why We Exist," webpage, undated-e. As of December 19, 2024: https://iteachforpakistan.org/why-we-exist

Teach For Pakistan, "Theory of Change for RAND's Study of Teach For Pakistan," March 20, 2023.

Teach For Pakistan, *Annual Report 2022–23*, Teach For All, 2024.

TFP—*See* Teach For Pakistan.

UNICEF—*See* United Nations International Children's Emergency Fund.

United Nations International Children's Emergency Fund, *Child Friendly Schools Manual*, 2009.

United Nations International Children's Emergency Fund, Pakistan, "Education," webpage, undated. As of December 19, 2024: https://www.unicef.org/pakistan/education

United Nations International Children's Emergency Fund, Pakistan, *National Nutrition Survey 2018: Key Findings Report*, Nutrition Wing, Ministry of National Health Services, Regulations and Coordination, Government of Pakistan, 2018.

Vossen, Helen G. M., Jessica T. Piotrowski, and Patti M. Valkenburg, "Development of the Adolescent Measure of Empathy and Sympathy (AMES)," *Personality and Individual Differences*, Vol. 74, February 2015.

West, Martin R., Katie Buckley, Sara Bartolino Krachman, and Noah Bookman, "Development and Implementation of Student Social-Emotional Surveys in the CORE Districts," *Journal of Applied Developmental Psychology*, Vol. 55, March–April 2018.

West, Martin R., Matthew A. Kraft, Amy S. Finn, Rebecca E. Martin, Angela L. Duckworth, Christopher F. O. Gabrieli, and John D. E. Gabrieli, "Promise and Paradox: Measuring Students' Non-Cognitive Skills and the Impact of Schooling," *Educational Evaluation and Policy Analysis*, Vol. 38, No. 1, March 2016.

White, Halbert, "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, Vol. 48, No. 4, 1980.

World Bank, "Five Major Challenges to Girls' Education in Pakistan," data visualization, June 2024. As of December 19, 2024: https://datatopics.worldbank.org/dataviz/girls-education-pakistan

Yang, Dongsheng, and Jarrod E. Dalton, "A Unified Approach to Measuring the Effect Size Between Two Groups Using SAS," *SAS Global Forum*, April 2012.

Yin, Robert K., *Qualitative Research from Start to Finish*, Guilford Publications, 2015.