Working Paper No. 600

The Impact of Teacher Professional Development Programs on Student Achievement in Rural China

Meichen Lu | Prashant Loyalka | Yaojiang Shi | Fang Chang

Chengfang Liu | Scott Rozelle

January 2017

Stanford Center for International Development

John A. and Cynthia Fry Gunn Building | 366 Galvez Street, Stanford, CA 94305-6015

The Impact of Teacher Professional Development Programs on Student Achievement in Rural China

Meichen Lu[a], Prashant Loyalka[b, c], Yaojiang Sh[a*], Fang Chang[a], Chengfang Liu[d] and Scott Rozelle[b]

[a]Center for Experimental Economics in Education, Shaanxi Normal University, 710119 Xi'an, Shaanxi, China

[b]Rural Education Action Program, Freeman Spogli Institute for International Studies, Stanford University, Stanford, CA 94305-6055, United States

[c]Graduate School of Education, Stanford University, Stanford, CA 94305-6055, United States

[d]School of Advanced Agricultural Sciences, Peking University, 100871 Beijing, China

Abstract: There is a significant gap in academic achievement between rural and urban students in China. Policymakers have sought to close this gap by improving the quality of teaching in rural areas through teacher professional development (PD) programs. Despite billions of dollars in investment, however, there is limited evidence on the effectiveness of such programs. In this paper, we evaluate the impact of a PD program-National Teacher Training Program (NTTP) on the academic achievement of students in rural China. We further examine the causal chain through which the program does or does not impact student achievement. By analyzing data on 84 teachers and 3,066 students from one Western province, we find that at a minimum, the NTTP has no effect on math achievement. In fact, it may even harm student academic achievement. We also find that while the program has a positive effect on math teaching knowledge of teachers, it has no significant effect on teaching practices in the classroom. Taken together, these results indicate that teachers may have improved their knowledge for teaching from NTTP, but did not apply what they learned to improve teaching practices or student learning.

Keywords: Teacher Professional Development; Teacher Knowledge; Student Achievement; Impact Evaluation; Rural China

***Corresponding author**
Yaojiang Shi
P.O. box 9, Shaanxi Normal University
620 Chang'an Rd West, Chang'an District, Xi'an City
Shaanxi Province, P.R.China, 710127.
E-mail: shiyaojiang7@gmail.com;
Tel.: +86 138 9283 3777; Fax: +86 029 8153 0872;

Acknowledgments: The authors would like to acknowledge the funding of the 111 project (高等学校学科创新引智计划资助, B16031).

The Impact of Teacher Professional Development Programs on Student Achievement in Rural China

Students in rural Chinese schools have less access to education than their urban peers; and what education they get is lower quality. Rural students achieve low levels of educational attainment and persistently drop out of school across all post-primary levels of schooling (Liu et al., 2009; Yi et al., 2012; Shi et al., 2015). Several empirical studies found that in rural junior high schools, the dropout rate ranges from 18 to 31 percent across a number of Chinese provinces (Li et al., 2015; Wang et al., 2015; Shi et al., 2015; Mo et al., 2013; Loyalka et al., 2013; Yi et al., 2012). In large cities, nearly 90 percent of students attend academic high school; in contrast, only 25 percent of students finish academic high school in rural areas (Liu et al., 2009). Meanwhile, in terms of academic achievement, students from rural areas perform significantly worse than urban students on challenges like competitive school entrance exams (Loyalka et al., 2014).

Many factors might be causing rural students to underperform their urban counterparts in educational outcomes. One possible reason for the achievement gap is fundamental differences in family investment in, and parental care for, students in rural versus urban areas (World Bank, 2001; Wang et al., 2009; Huang and Du, 2007; Ye and Pang, 2011; Luo et al., 2009). Research from developed countries has shown quality of teaching to be another important factor that may affect educational outcomes among both advantaged and disadvantaged students (Darling-Hammond, 2000; Rivkin, Hanushek and Kain, 2005; UNESCO, 2006). In a survey of education polices and student achievement that was conducted in all fifty of the United States, Darling-Hammond (2006) reported that teacher preparation and certification significantly correlates with student academic performance.

Such work highlights the importance of teaching quality in improving student academic performance.

There are fewer studies of teacher quality in developing countries, but those available confirm that differences in teacher quality can significantly impact student achievement. One such study found that in Peru, teachers with high achievement in math increased student achievement on standardized math tests by about 9 percent of a standard deviation (Metzler and Woessmann, 2012). In China, teachers of the highest professional rank more positively affect rural students' achievement than teachers of lower rank (Chu et al, 2015).

While there may be a number of different ways to improve the quality of teaching for rural students (for example, improving incentives for teachers—Loyalka et al., 2015; Muralidharan and Sundararaman, 2011; Muralidharan, 2012), policymakers in developing countries have placed great stock in teacher professional development (PD) programs (Vegas, 2007; Villegas-Reimers, 2003; Cobb, 1999). In theory, PD programs seek to help teachers gain subject-specific knowledge (Dadds, 2001), use appropriate pedagogical practices (Darling-Hammond and McLaughlin, 1995; Schifter et al., 1999), develop positive attitudes towards teaching (Cobb, 2000), and ultimately improve the learning of students (Villegas-Reimers, 2003). The importance of PD is further supported by empirical evidence from developed countries. For example, in a review of experimental evaluations of PD programs in developed countries, Yoon et al. (2007) found that the academic achievement of students whose teachers participated in PD programs increased by 0.54 standard deviations compared to students whose teachers did not participate in PD. These positive findings lend credence to the efforts of policymakers from developing countries who are in favor of using PD to improve the quality of teaching in their rural areas.

In recent years the Chinese government has invested heavily in teacher PD programs. In 2010, China's government launched the National Teacher Training Program (NTTP), the country's flagship teacher PD program (MOE and MOF, 2010).[1] Beyond improving teaching quality in general, one of the major goals of the NTTP is to improve the quality of teaching in rural regions and improve the learning of rural students (MOE and MOF, 2010). Given the high level of investment in the NTTP, and given its ambitious goals, the program is currently one of the key national government initiatives for improving human capital of rural students and improving the equity of educational outcomes between rural and urban students in China. In order to improve the PD of teachers, the Ministry of Education has prescribed that the training content should focus on ethics in teaching, subject-specific knowledge, and pedagogical practices in proportions of 10%, 40%, and 50%, respectively (MOE, 2012).

Although policymakers in China and other developing countries invest billions of dollars in teacher PD programs each year (e.g. Yan, 2013; Government of Chile, 2003; Government of India, 2013), there is only limited evidence of whether these programs are effective (Bruns and Luque, 2014). At best, policymakers have only obtained subjective feedback from the teachers that have been trained and the trainers running the programs (e.g. Zuo and Su, 2012). Although researchers also have compared different PD programs, studied the potential of new teaching technologies, and assessed teacher learning and other outcomes within these programs using more quantitative approaches (Olakulehin, 2007; Garet et al.,

[1] Beyond the NTTP, there are many other teacher PD programs that are run by local governments. As the nation's flagship program, the NTTP involves much larger expenditures per teacher and more prestige for participation than local teacher PD programs can offer.

2001; Borko, 2004; Overbaugh and Lu, 2008; Owston et al., 2008), few studies have demonstrated a causal relationship between the evaluated programs and student outcomes. In particular, there have been few large-scale, rigorous evaluations of teacher PD programs on student achievement in developing countries. In fact, one of the only exceptions in a developing country context that we know of is Yoshikawa et al. (2015)'s study, which assesses the impacts of a pilot PD program for early childhood education teachers in 64 preschools in Chile. This study found that the PD program moderately impacted emotional and instructional support as well as classroom organization. The study's results, however, do not indicate any impact on the cognitive outcomes of students. To our knowledge, there has been no large-scale empirical evaluation of PD programs in China. Only one recent study describes the current teacher PD program system in China (Liu et al., 2016); however, the study fails to evaluate the impact of PD programs on student achievement.

The overall goal of this study is to examine the effect that the NTTP has on student achievement in rural China. To reach this goal, we first examine the impact of the NTTP on the math achievement of rural students. Next, we seek to understand the causal mechanism through which the NTTP is or is not able to affect student achievement. To do this, we will examine whether the NTTP was able to improve either the knowledge of math or the teaching practices of teachers. We will also seek to come up with an explanation of why the NTTP might have a positive (or zero or negative) effect on students.

The remainder of the paper is organized as follows. In section 2, we describe the conceptual framework for how a teacher PD program might be expected to improve student achievement. In Section 3, we describe the study's methodology, including the sampling process, the nature of the intervention, our data collection effort, and our analytical approach. Section 4 reports and discusses the results of the study. Section 5 concludes.

2. Conceptual Framework

Based on a review of literature on teacher PD programs (Chapman et al., 2000; Cohen & Hill, 2000; Garet et al., 2001; Fishman et al., 2003; Yoon et al., 2007), in this section we describe a conceptual framework of the causal chain in which a PD program (here, the NTTP) ultimately raises student achievement (Figure 1). This framework assumes that teachers who attend the PD program will actually learn the materials taught during the program, and that the material is designed such that teachers' instruction approaches will improve.

Figure 1 presents our framework, a three-step process through which teacher PD programs improve student achievement. First, teachers who receive training spots must actually attend the PD program and fully participate in the program without missing any of the training. Second, the PD programs need to enhance the teaching knowledge and/or skills of participating teachers. Third, after learning the material that was presented in the training program, teachers need to apply this knowledge within their classrooms. However, in order to apply the knowledge to their teaching practice, teachers must have adequate motivation, beliefs, and skills to overcome such barriers as limited time for preparation and instruction, or limited materials and resources (Borko, 2004; Showers, Joyce and Bennett, 1987). Therefore, for a PD program to end up being able to improve student achievement, all three steps need to be fulfilled. If one link in this causal chain is weak or missing, improved student outcomes are less likely to be achieved.

3. Method

3.1 Sampling

In this study, we conducted an evaluation of the NTTP for primary school math teachers in Shaanxi Province. Shaanxi Province is located in northwest China. Its provincial GDP per capita is 46,928 yuan (approximately 7640 dollars); Shaanxi ranks fourteenth among the 31 mainland provinces of China in terms of GDP per capita (CNBS, 2014). There are ten prefectures in Shaanxi Province. Our study took place in two of them: Weinan, which ranks ninth in terms of GDP per capita among all Shaanxi prefectures; and Xianyang, which ranks fifth (Shaanxi Statistical Bureau, 2014).

Teachers who participated in the evaluation were selected from a much larger group of teachers who had been chosen to receive NTTP training through a standard process across China. According to China's national teaching development policy, teachers are required to participate in a certain number of hours of PD training each year. However, as the policy rules are written, there is no reference to which type of program counts towards these hours. China's government offers non-NTTP PD programs at different levels – provincial, prefectural and county (Liu et al., 2016). In the case of the (national) NTTP, the central office in Beijing decides which trainings will take place each year and then each province's NTTP office allocates training slots to each prefecture. Each prefecture allocates the slots to each county. Finally, at the county education bureau level, officials choose the schools that are required to send teachers to that year's NTTP. The criteria used by county officials differ across regions, but during interviews we learned that schools were often given priority according size and past rates of participation in the program. The exact registration process for individual teachers varies across schools.

To select our treatment sample for the evaluation, the research team first secured access to the list of all teachers who had been enrolled in the NTTP program (for the year of our study). It should be noted that the research team was not involved in choosing the NTTP teachers; they had already been chosen by the time we launched our study. In the study prefectures, a total of 63 teachers were on the initial list of NTTP trainees.

The next step of the sampling process was to choose the actual treatment teachers from the overall list of trainees. In order to select a sample representative of the majority of schools in Xianyang and Weinan prefectures, we excluded teachers who came from schools that did not provide all six grades of primary school education. We further limited our sample to teachers who taught math in one of our sample grade levels (3^{rd}-6^{th} grade). After applying the exclusion criteria, we were left with 34 treatment teachers across 34 treatment schools.

The final step of choosing the sample in the treatment schools was to choose the exact class and set of students that were supposed to be the beneficiaries of the NTTP. If a teacher in our sample taught math at more than one grade level, we randomly selected one grade level for inclusion in our study. If a teacher taught more than one math class in the same grade, we included all classes in that grade level in our sample.

The next step of our sampling procedure was to select teachers that did not participate in the NTTP to serve as a control group. Our control group was made up of teachers selected in one of two ways. The first group of control teachers was selected from the same schools as the treatment teachers (*within-school control teachers*). The second group of control teachers was selected from different schools from the treatment teachers (*across-school control teachers*).

The within-school control teachers were selected using the following procedure: We called each treatment school to ask whether there were other math teachers teaching the same grade level as our treatment teacher. If there was more than one other teacher, one was randomly selected as our control teacher. (If there was only one, that teacher was the control.) If there was not another math teacher in that grade level, then we did not select a within-school control teacher from that school. In total, we were able to sample 16 within-school control teachers. For each within-school control teacher who taught more than one math class at the same grade level as our treatment teacher, we randomly selected one of those classes for inclusion in our sample.

The across-school control teachers were selected according to the following protocol: First, we travelled to the Bureau of Education of each county where our treatment schools were located and collected information on all primary schools in that county. Specifically, we collected information on: a) the distance between our treatment school and the other primary schools in that county; b) the number of students in each school; and c) the *tongkao* (standard exams organized by the Bureau of Education of each county) ranking of each school. For each treatment school, we chose a control school in that county that best matched the treatment school on these three criteria (closest in terms of distance; similar number of students; similar tongkao ranking). In total, there were 34 pairs of treatment and control schools.

After we selected the control schools, we used the following two-step procedure to select across-school control teachers: First, we contacted the principal of each school and asked for a list of math teachers that taught the same grade level as the treatment teacher from the matched school. Next, we randomly selected one teacher from that grade level for each control school. In total, we selected 34 across-school control teachers. If the teacher taught more than one math class in the same grade as our treatment teacher, we randomly selected one class for inclusion in our sample.

The sample at the time of the baseline survey (October 2014) consisted of 84 teachers (34 treatment teachers; 16 within-school control teachers; and 34 across-school control teachers) and 3289 students. By the time of the endline survey in February 2015, however, there was some student attrition. Due to various reasons (e.g. dropouts, absences, death, and missing data), 223 students (71 students in the treatment group and 152 students in control group) did not complete our endline survey. This means that the attrition rate was 7 percent. There was no attrition of teachers. Hence, in total, our final sample included 84 teachers and 3066 students. In total, there were 1141 students in the treatment schools and 1922 students in control schools.

To examine if attrition of students affected our results, we regressed attrition status on the treatment variables. Comparing the attrition rates between the treatment group and the control group, the results show that treatment status did not affect the attrition rate (Table 1). Furthermore, we checked for balance on observable characteristics among our students and teachers who participated in both the baseline and endline surveys. According to Table 2, we find that treatment and control students are fully balanced (rows 1 to 7). In terms of teacher characteristics, Table 2 shows that one key outcome variable (standardized teacher math teaching knowledge test scores—row 8) and most control variables (including teacher gender, whether teacher completed university, and whether teacher has highest rank–rows 9-11) are balanced. We did find significant differences in terms of two variables: whether the teacher majored in math and his/her teaching experience (Table 2, rows 12 to13)[2]. To account for the imbalance of these variables in our subsequent analyses, we controlled for them in our regressions. After adding control variables, we saw no significant change in the point estimates for each outcome, suggesting that the imbalance is not significant enough to affect internal validity.

3.2 Intervention

Our intervention was conducted at the teacher-level. All treatment teachers participated in the NTTP after the baseline survey. The PD program they participated in was organized by two top-ranked primary schools under the oversight of the Shaanxi provincial NTTP office. The research team did not have any input into the training. This was a training program run fully by the education system. The instructors in the program were fully blind to the study. During the two-week program, all treatment teachers were released from regular teaching duties to receive training at a centralized location in the prefecture.

The curriculum of the treatment teachers' PD program followed the basic mandated framework of the NTTP. According to the teaching materials that we observed the treatment teachers using, there were sections of the training classes that were focused on ethics in teaching, subject-specific knowledge, and pedagogical practices (MOE, 2012). The instructors of the NTTP were top teachers from the highest-ranked primary schools in the prefecture. These teachers were selected as trainers because of their reputation for being knowledgeable about teaching math. Training sessions were conducted for six hours per day;

[2] As it turns out none of the treatment teachers who attended the math NTTP were not math teachers. We had assumed that only math teachers would be assigned to a math PD program and so we choose math teachers to fill all of the slots in the control group.

sessions consisted primarily of lectures with short intervals for lesson modeling, questions, and discussion.

Beyond the basic components of the in-program training, the NTTP did not provide any in-classroom follow-up instruction. In other words, after the training, neither the NTTP instructors or any other NTTP staff went to any of the schools of treatment teachers. There was no attempt to observe whether teachers applied what they learned, to identify any weaknesses in the teaching practices of teachers, and/or to instruct teachers on how to address such weaknesses. The only post-training follow-up was through the establishment of an online group-chat for teachers (which was actually organized by the teachers themselves, not by the NTTP staff).

3.3 Data Collection

We collected data from student and teacher responses to a baseline survey (October 2014) and an endline survey (February 2015). The student baseline survey consisted of three blocks. In the first block, students provided basic demographic and family background information, including student age, gender, whether a student was a left-behind child,[3] boarding status, and whether his/her father or mother had completed at least junior high school.

In a second block, students filled out a checklist of practices they observed in their teachers that might have been influenced by participation in the PD program. This checklist included items such as whether their teacher clarified learning goals in each math class; provided enough time to let students express themselves; asked questions to ensure that

[3] A left-behind child is one whose parents have both left home, usually to work in an. These children are typically raised by a close relative- for example, grandparent (Zhou et al., .2015; Ye et al., 2006).

students understood lesson content; allowed students to discuss the new material being taught in groups; and/or provided different homework assignments according to the academic achievement of students. With this information, we created an index of teaching practices using the GLS weighting procedure described in Anderson (2008).

In the third block of the student survey, students were given a 30-minute standardized math test. The math test items were drawn from the Chinese National Curriculum Framework (MOE, 2011). We prepared and administered the test ourselves to ensure that students and teachers could not prepare for the test. The enumeration team closely proctored the exams in order to strictly enforce time limits and minimize cheating. The scores were standardized by scaling them into z-scores by subtracting the mean and dividing by the standard deviation (SD) of the math score distribution of all students tested. These standardized scores are used as our key measure for assessing student baseline math achievement.

We administered the teacher baseline survey in two blocks. In the first block, enumerators collected information on the characteristics of teachers. Specifically, we solicited information on each teacher's gender, educational attainment, whether he/she majored in math, whether he/she had obtained the highest teaching rank, and his/her teaching experience.

In the second block, we administered the 40-minute Learning Mathematics for Teaching (LMT) test designed by a research team at the University of Michigan (Study of Instructional Improvement and Learning Mathematics for Teaching Project—Hill, Schilling, and Ball, 2004). The purpose of this test was to assess teacher knowledge of math instruction. The test not only evaluates whether teachers can answer the mathematics problems they assign students, but also assesses how well teachers are able to solve mathematical tasks that arise in the process of teaching (Hill, Schilling, and Ball, 2004). The LMT test has previously been used in many studies (Faulkner and Cain, 2013; Copur-Gencturk and Lubienski, 2013; Bell et al., 2010; Agodini et al., 2009; Delaney et al., 2008; Hill, Rowan, and Ball, 2005). The teacher test scores were standardized for ease of interpretation.

In February 2015, we returned to all sample schools to conduct a follow-up survey. The procedures for conducting the endline survey were identical to those for the baseline survey with two exceptions. First, we did not ask students or teachers for their basic background information. Second, we gave teachers a different version of the LMT than they took during the baseline survey.

3.4 Analytical Approach

In this subsection we introduce the analytical approach we used to examine the impact of the NTTP on both student academic achievement, and on teachers' knowledge of math teaching practices. We assess the impacts using ordinary least squares (OLS) regression with school-pair fixed effects.

3.4.1 OLS Model for examining the impact of the NTTP on student academic achievement

We use unadjusted and adjusted OLS regression analysis to estimate how student achievement in the treatment group changed relative to student achievement in the control group. The unadjusted model is:

$$Y_i = \alpha_0 + \alpha_1 T_i + \varphi_p + \varepsilon_i \qquad (1)$$

where Y_i represents the endline math achievement score of student *i*; T_i is a dummy variable that equals 1 if the student participated in the NTTP treatment and 0 otherwise; and ε_i is a random error term. It also includes a school-pair fixed effect, φ_p. Note that since there were

34 pairs of treatment schools and control schools in our sample, the matrix φ_p is made up of 34 school pair dummy variables.

To control for potential confounding effects of student and teacher characteristics, we also adjusted for additional covariates (X_i and Z_i). We call equation (2) below our adjusted school-pair fixed effect model:

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 X_i + \beta_3 Z_i + \varphi_p + \varepsilon_i \qquad (2)$$

where the X_i represents a vector of student characteristics at baseline, including *standardized baseline math test score*, *age, female student (1=yes, 0=no),* student is a *left-behind child (1=yes, 0=no)*, *student lives at school (1=yes, 0=no), mother completed junior high school (1=yes, 0=no)* and *father completed junior high school (1=yes, 0=no).* The additional term Z_i represents a vector of math teacher characteristics, including baseline *female teacher (1=yes, 0=no),* whether the teacher *completed university (1=yes, 0=no),* whether the teacher *majored in math (1=yes, 0=no)*, whether the teacher has the *highest rank (1=yes, 0=no)*, and *teaching experience (1=over 15 years, 0=equal or below 15 years)*. See Table 2 for descriptive statistics of these variables. In addition, equation (2) also includes a school-pair fixed effect term φ_p.

To examine whether the NTTP had greater impacts on some subgroups than others, we used a heterogeneous effects model to estimate treatment parameters. The heterogeneous effects model is essentially equation (2) with an additional interaction term between the NTTP treatment variable and a student background variable (measured using the baseline survey data). We chose several different types of student background variables to interact with the treatment variable, including *baseline student math test score*, *student age*, *female* *student* and *student is a left-behind child.* In all regressions, we included a school-pair fixed effect.

3.4.2 OLS Model for examining the impact of the NTTP on math teaching knowledge and teaching practices

The primary outcome variables for teachers were (a) math teaching knowledge test score, and (b) the teaching practices index. As with examining the impacts of the NTTP on student outcomes, we first conducted unadjusted analyses using the following model:

$$Y_j = \eta_0 + \eta_1 T_i + \varphi_p + \varepsilon_i \qquad (3)$$

where Y_i represents the outcome variable of teacher *j*; T_i is a dummy variable that equals 1 if the teacher participated in the NTTP and 0 otherwise; and φ_p is a school-pair fixed effect.

We also conducted an adjusted analysis, which controls for teacher-level covariates (Z_i). The adjusted school-pair fixed effect model is as follows:

$$Y_j = \lambda_0 + \lambda_1 T_j + \lambda_2 Z_i + \varphi_p + \varepsilon_j \qquad (4)$$

where term Z_i represents a vector of math teacher characteristics, including *baseline math teaching knowledge test score/baseline math teaching practice index,* teacher *gender*, whether the teacher *completed university,* whether the teacher *majored in math*, whether the teacher has the *highest rank*, and *teaching experience*.

4. Results

4.1 Effect of the NTTP on student academic achievement

Our analyses suggested that the NTTP has no impact on student academic achievement. In the unadjusted model, the estimate of the impact of the NTTP on student achievement was small in magnitude and not statistically significant (Table 3, column 1, row 1). After adding controls for student and teacher characteristics and school fixed effects, we found that the NTTP actually produces a small, negative impact on student academic achievement. The result was not significantly different from zero, however. Finally, results from the adjusted school-pair fixed effect model showed that the NTTP decreased the academic achievement of treatment students by 0.07 standard deviations (Table 3, column 2, row 1). This result is statistically significant at the 5% level. In other words, at a minimum, the evidence suggests that teacher participation in the NTTP does not increase student achievement. In fact, it may even harm student achievement. Given the government's sizeable investments in the NTTP, these results are concerning. Improving the achievement of rural students is one of the top goals of the nation's flagship teacher PD program.[4]

We were also interested in determining whether the NTTP has differential impacts on different subgroups of students. To test this, we examined heterogeneous effects by applying the treatment variable separately with each student background characteristic of interest (baseline student math test score, student age, student gender， and whether the student is a left-behind child). This analysis showed no significant heterogeneous impacts of the NTTP on student academic achievement for any of these variables (Table 4, rows 2 to 5).

4.2 Unpacking the causal chain: effect of the NTTP on teacher math teaching knowledge and teaching practices

Our analysis showed that the NTTP did not improve student academic achievement (and may even be harmful for student learning). Why might this be the case? It is possible that the NTTP was simply ineffective in teaching teachers new skills. To check for this

[4]In addition, we ran another analysis that considers the two types of control teachers (within-school control teachers and across-school control teachers). This analysis was carried out by adding a dummy variable that equals 1 if the teacher was a within-school control teacher and 0 otherwise. The results of the NTTP on student academic achievement using this approach are substantively identical to those that we report in the manuscript. These results are reported in a supplemental appendix available online: (https://reap.fsi.stanford.edu/publication/impact-teacher-professional-development-programs-student-achievement-rural-china).

possibility, we looked at the impact of the NTTP on teacher math teaching knowledge. According to our unadjusted model, the NTTP increased the math teaching knowledge test scores of treatment teachers by 0.55 standard deviations relative to control teachers (Table 5, column 1, row 1). This result is significant at the 5% level. The results stayed consistent after controlling for school-pair fixed effects and teacher characteristics. We found treatment teachers' average math teaching knowledge test score to be 0.42 standard deviations higher than that of control teachers, which is significant at the 5% level (Table 5, column 2, row 1).

Altogether, though the NTTP failed to improve student achievement, our results tentatively suggest that the teachers may have learned more about how to teach math. This finding suggests that the null impact of the NTTP on student achievement cannot be attributed to a lack of improvement in the math teaching knowledge. However, it is clear that whatever knowledge teachers picked up in the training was not translated into improved student achievement.

One possible explanation for the disconnect between teacher and student achievement is that even if the NTTP improved teacher knowledge of teaching math, teachers may not have changed their teaching practices after completing the NTTP. Indeed, even though the Ministry of Education has prescribed that 50% of NTTP training material focus on pedagogical practices, our results show that the effect of the NTTP on teaching practices is negligible. Although the unadjusted model shows that the teaching practice index of the average teacher in the treatment group is 0.03 standard deviations higher than that of the average teacher in the control group, this is not significant (Table 6, column 1, row 1). The results from our adjusted school-pair fixed effect model also show that the NTTP has no significant impact on teaching practices (Table 6, column 2, row 1). In light of the possibility that student responses are unreliable (and to carry out a supplemental approach), we conducted independent classroom observations in 8 randomly selected classrooms (4 control classrooms and 4 treatment classrooms)[5]. We did not observe any difference in teaching styles between treatment and control teachers.

These results indicate that treatment teachers did not use the teaching knowledge they learned in the NTTP to improve their teaching practices. Although it was not possible to measure this quantitatively, there are two reasons why this may have happened. One possibility is that the teaching knowledge that was taught by the NTTP may not have been useful or appropriate for use in the classroom. In fact, teachers (whom we interviewed during pretests/piloting) often complained that much of the material they learn in NTTP sessions is not relevant. Another possibility is that teachers may have lacked the ability or incentives to convert their increased knowledge into improved teaching practices and ultimately into student achievement.

---

5 As part of our effort to better understand whether teachers apply their knowledge to their teaching practice, we conducted independent classroom observations immediately after the endline survey (February 2015). Our classroom observation approach was based on a modified protocol of the Classroom Assessment Scoring System (CLASS). The CLASS protocol provides a common framework to observe quality of classrooms across grades (Pianta et al., 2007; Pianta et al., 2008). CLASS was designed to observe and document the instructional support, emotional support, classroom organization and student engagement in the classroom (Pianta et al., 2007; Pianta et al., 2008). In our carrying out of CLASS-based observations in our study schools, a team of 2 enumerators conducted the observations in 8 randomly selected classrooms (4 treatment classrooms and 4 control classrooms). Before conducting the observations, the enumerators participated in a step-by-step instructional training course on how to conduct classroom observations. When enumerators conducted the observations, they sat in the back of the classroom, observed and recorded the classroom practice in terms of certain behavioral, emotional and physical markers without any intervention into or disturbance of the class. During the observation period, the two enumerators independently recorded their individual observations without communicating with each other. After each class, they immediately compared their records and decided on the final observation results.

There may also have been offsetting effects. In particular, when the treatment group teachers left the classroom for two weeks to participate in the NTTP, their absence from the classroom may have disrupted student learning. The NTTP evaluated in this study was conducted during the semester in urban areas far from the home schools of our sample teachers. As a result, teachers had to leave their classrooms for two weeks to participate in the training. In most rural schools, each grade only has one math teacher. Therefore, the absence of the trained teacher may result in the substitution of teachers who are less familiar with the course curriculum or less motivated to teach well into the treatment teacher's classroom for the duration of the NTTP. It is possible that this temporary disruption may have had a negative effect on student learning that offset or even exceeded any positive effect of the NTTP on teaching quality.

5. Conclusions

In general, we find that at a minimum, teacher participation in the NTTP had no positive impact on student achievement. We did not discover a statistically significant difference in math achievement between treatment and control students in any of our models. In fact, teacher participation in the NTTP may have even lowered student academic achievement. Our findings also demonstrate that there were no significant heterogeneous effects.

Why did the NTTP not yield positive impacts? According to our analysis, we can tentatively conclude that the lack of impact on student outcomes was not due to a lack of learning among the trained teachers. We find evidence to suggest that teachers who received NTTP training did have improved math teaching knowledge relative to control teachers.

However, this improvement in knowledge did not result in any significant effect on teaching practices. If the NTTP is faulty in its design, it is likely because it fails to induce teachers to apply what they learned in the training within their classrooms. In addition, having their teachers attend the NTTP during the semester may have disrupted normal learning for the students. This mechanism may help to explain the negative impact of the NTTP on student academic achievement.

It should be noted that the study may suffer from a potential limitation because we only examined changes in student achievement after one semester. It could be that the change in teacher knowledge eventually will lead to greater gains for students over the course of a school year. Also, it could be that future cohorts of students benefit from the change in teacher knowledge. These hypothesized effects would nevertheless have to rely on the teacher having made a permanent change in the way they were teaching. Given the fact that teachers did not change their math teaching practices in the period of time immediately after participating in the study, the possibility of students benefiting from a change in teacher knowledge may be negligible.

The results of this study contribute to a broader policy debate about how to effectively invest in teacher PD programs, especially in rural China. Recently, there has been increasing support from officials in the Ministry of Finance of the People's Republic of China for greater investment in national teacher PD programs for rural teachers (MOE and MOF, 2015). Our results suggest that if the government is interested in providing rural teachers with more opportunities to participate in the NTTP, the programs may need to place a greater emphasis on helping teachers convert new skills and knowledge acquired during training into actual results in the classroom.

There are programs in other countries that do this. For example, “clinical supervision,” has been used as a method to foster teacher development through discussion, observation, and analysis of teaching “in the clinic of the classroom” (Grimment and Crehan, 1992). This type of approach has been used in several countries, including Pakistan (Gardner, 1995), Nigeria (Tatto, 1997), Israel (Barak et al., 1997), and Brunei (Bourke, 2001). More research is needed to determine the effectiveness of such programs. Similar classroom-based interventions, however, could offer the potential to improve teaching quality and student educational outcomes in China.

References:

Agodini, R., Harris, B., Atkins-Burnett, S., Heaviside, S., Novak, T., and Murphy, R., 2009. Achievement Effects of Four Early Elementary School Math Curricula: Findings from First Graders in 39 Schools. NCEE 2009-4052. National Center for Education Evaluation and Regional Assistance.

Anderson, M., 2008. Multiple Inference and Gender Differences in The Effects of Early Intervention: A Revaluation of the Abecedarian, Perry Preschool, and Early Training Projects. Journal of the American statistical Association. 103. 1481-1495.

Barak, M., Pearlman-Avnion, S., and Glanz, J., 1997. Using Developmental Supervision to Improve Science and Technology Instruction in Israel. Journal of Curriculum and Supervision, 12(4), 367.

Bell, A., Wilson, M., Higgins, T., and McCoach, B., 2010. Measuring the Effects of Professional Development on Teacher Knowledge: The Case of Developing Mathematical Ideas. Journal for Research in Mathematics Education, 479-512.

Borko, H., 2004. Professional Development and Teacher Learning: Mapping the Terrain. Educational Researcher, 33(8), 3-15.

Bourke, M., 2001. The Role of the TP [Teaching Practice] TESL [Teaching English as a Second Language] Supervisor. Journal of Education for Teaching: International Research and Pedagogy, 27(1), 63-73.

Birgin, A. , Braslavsky , C., and Duschatzky , S., 1992. The Training of Teachers: Towards the Construction of A New Paradigm for Processing. Braslavsky, C. and Birgin, A. (Eds.), Teacher Training, Impact, Past and Present. Buenos Aires: Miño and Davila . (In Spanish)

Bruns, B., Luque, J., 2014. Great Teachers: How to Raise Student Learning in Latin America and the Caribbean. Washington, DC: The World Bank.

Castro, E., 1991. Teacher Training in Latin America: A Challenge that Requires Response. Santiago, Chile: OREALC/ UNESCO. (In Spanish)

China National Bureau of Statistics (CNBS)., 2014. China Statistical Yearbook, China Statistics Press, Beijing.

Chu, H., Loyalka, P., Chu, J., Qu, Q., Shi, Y., Li, G., and Rozelle, S., 2015. The Impact of Teacher Credentials on Student Achievement in China. China Economic Review, 36, 14-24.

Cobb, L., 1999. An International Comparison of Teacher Education. ERIC Digest. Washington DC: ERIC Clearinghouse on Teaching and Teacher Education.

Cobb, J. 2000. The Impact of a Professional Development School on Preservice Teacher Preparation, Inservice Teachers' Professionalism, and Children's Achievement: Perceptions of Inservice Teachers. Action in Teacher Education, 22(3), 64-76.

Copur-Gencturk, Y., Lubienski, T., 2013. Measuring Mathematical Knowledge for Teaching: A Longitudinal Study Using Two Measures. Journal of Mathematics Teacher Education, 16(3), 211-236.

Dadds, M., 2001. Continuing Professional Development: Nurturing the Expert Within. Soler, J., Craft, A., Burgess, H., (Eds.), Teacher Development: Exploring Our Own Practice. London: Paul Chapman Publishing and The Open University.

Darling-Hammond, L., McLaughlin, W., 1995. Policies that Support Professional Development in An Era of Reform. Phi delta kappan, 76(8), 597.

Darling-Hammond, L., 2000. Teacher Quality and Student Achievement: A review of State Policy 2000. Education Policy Analysis, 8, 1.

Delaney, S., Ball, L., Hill, C., Schilling, G., and Zopf, D., 2008. Mathematical Knowledge for Teaching: Adapting US Measures for Use in Ireland. Journal of Mathematics Teacher Education, 11(3), 171-197.

Faulkner, N., Cain, R., 2013. Improving the Mathematical Content Knowledge of General and Special Educators Evaluating a Professional Development Module That Focuses on Number Sense. Teacher Education and Special Education. The Journal of the Teacher Education Division of the Council for Exceptional Children, 36(2), 115-131.

Fishman, J., Marx, W., Best, S., Tal, T., 2003. Linking teacher and student learning to improve professional development in systemic reform. Teaching and Teacher Education, 19(6), 643–658.

Garder, R., 1995. On-service Teacher Education. Anderson, L., (Ed.), International Encyclopedia of Teaching and Teacher Education (Second Edition). London: Pergamon Press.

Garet, S., Porter, C., Desimone, L., Birman, F., and Yoon, S., 2001. What Makes Professional Development Effective? Results from a National Sample of Teachers. American Educational Research Journal, 38(4), 915-945.

Chapman, W., Chen, X., Postiglione, A., 2000. Is Preservice Teacher Training Worth the Money? A Study of Teachers in Ethnic Minority Regions of the People's Republic of China. Comparative Education Review, 44(3), 300-328.

Cohen, K., Hill, C., 2000. Instructional policy and classroom performance: The mathematics reform in California. Teachers College Record, 102(2), 294–343.

Government of Chile. 2003. Attracting, Developing, and Retaining Effective Teachers: OECD Activity, Country Background Report for Chile. Santiago: http://www.oecd.org/chile/26742861.pdf

Government of India .2013. Innovative Teacher Training Modules. New Delhi: http://www.teindia.nic.in/Te_Trg_Module.aspx

Grimmett, P., Crehan, P., 1992. The Nature of Collegiality in Teacher Development: The Case of Clinical Supervision. Teacher Development and Educational Change, 12(2), 56-85..

Hill, C., Schilling, G., and Ball, L., 2004. Developing Measures of Teachers' Mathematics Knowledge for Teaching. Elementary School Journal, 105, 11–30.

Hill, C., Rowan, B., and Ball, L., 2005. Effects of Teachers' Mathematical Knowledge for Teaching on Student Achievement. American Educational Research Journal, 42(2), 371-406.

Huang A. Du, X., 2007. Comparative Analysis of Urban and Rural Difference in Family Education in China. Journal of Yibing University, 1, 107-110. (In Mandarin Chinese)

Li, F., Loyalka, P., Liu, C., Yi, H., Zhang, L., Chu, J., Johnson, N. and Rozelle, S., 2015, The Impact of Conditional Cash Transfers on the Matriculation of Junior High School Students into Rural China's High Schools. Working Paper, The Rural Education Action Program, Freeman Spogli Institute for Foreign Policy, Stanford, CA. Available at SSRN 2610584

Liu, C., Zhang, L., Luo R., Rozelle, S., Sharbono, B., Shi, Y., 2009. Development Challenges, Tuition Barriers and High School Education in China. Asia Pacific Journal of Education, 29(4), 503–520

Liu, H., Liu, C., C, F., Loyalka, P., 2016. Describing the Landscape of Teacher Training in China. China & World Economy, 24(3), 86-104.

Loyalka, P., Liu, C., Song, Y., Yi, H., Huang, X., Wei, J., and Rozelle, S., 2013. Can Information and Counseling Help Students from Poor Rural Areas Go to High School? Evidence from China. Journal of Comparative Economics, 41(4), 1012-1025.

Loyalka, P., Chu, J., Wei, J., Johnson, N., Reniker, J., and Rozelle, S., 2014 Inequalities in the Pathway to College in China: When Do Students from Poor Areas Fall Behind? Working Paper, The Rural Education Action Program, Freeman Spogli Institute for Foreign Policy, Stanford, CA.

Luo, R., Shi, Y., Zhang, L., Liu, C., Rozelle, S., and Sharbono, B., 2009. Malnutrition in China's Rural Boarding Schools: the Case of Primary Schools in Shaanxi Province. Asia Pacific Journal of Education, 29(4), 481-501.

Metzler, J., Woessmann, L., 2012. The Impact of Teacher Subject Knowledge on Student Achievement: Evidence from Within-teacher Within-student Variation. Journal of Development Economics, 99(2), 486-496.

Ministry of Education and Ministry of Finance (MOE and MOF). 2010. Notice of the Implementation of the National Training Plan for Primary and Secondary Education

Teachers. Government of China：
http://www.gov.cn/zwgk/2010-06/30/content_1642031.htm

Ministry of Education and Ministry of Finance (MOE and MOF). 2015. Notice of Reforming the National Training Plan for Pre-school, Primary and Secondary Education Teachers. Government of China：
http://www.mof.gov.cn/zhengwuxinxi/zhengcefabu/201509/t20150907_1452338.htm

Ministry of Education (MOE). 2011. Chinese National Curriculum Framework Accessed September 2, 2014, Government of China:
http://www.moe.gov.cn/publicfiles/business/htmlfiles/moe/s8001/201404/xxgk_1673 4 0.html

Ministry of Education (MOE). 2012. Training Framework of National Teacher Training Program. Government of China:
http://old.moe.gov.cn/ewebeditor/uploadfile/2012/10/12/20121012170215824.pdf

Mo, D., Zhang, L., Yi, H., Luo, R., Brinton, S., 2013. School Dropouts and Conditional Cash Transfers: Evidence from a Randomized Controlled Trial in Rural China's Junior High Schools. Journal of Development Studies, 49(2), 190-207.

Muralidharan, K., Sundararaman, V., 2011. Teacher Performance Pay: Experimental Evidence from India. The Journal of Political Economy, 119(1), 39–77.

Muralidharan, K., 2012. Long-Term Effects of Teacher Performance Pay: Experimental Evidence from India. Society for Research on Educational Effectiveness.

Olakulehin, K., 2007. Information and Communication Technologies in Teacher Training and Professional Development in Nigeria. Turkish Online Journal of Distance Education, 8(1), 133-142.

Overbaugh, R., Lu, R., 2008. The Impact of a NCLB-EETT Funded Professional Development Program on Teacher Self-efficacy and Resultant Implementation. Journal of Research on Technology in Education, 41(1), 43-61.

Owston, R., Wideman, H., Murphy, J., and Lupshenyuk, D., 2008. Blended Teacher Professional Development: A Synthesis of Three Program Evaluations. The Internet and Higher Education, 11(3), 201-210.

Pianta, C., Hamre, K., Haynes, J., Mintz, L., La Paro, M., 2007. Classroom Assessment Scoring System Manual, Middle/Secondary Version. Charlottesville: University of Virginia.

Pianta, C., La Paro, M., Hamre, K., 2008. Classroom Assessment Scoring System, Manual, K-3. Baltimore, MD: Brookes.

Rivkin, S., Hanushek, E., and Kain, J. 2005, Teachers, Schools, and Academic Achievement. Econometrica*,* 73(2), 417-458.

Schifter, D., Russell, S.J., Bastable, V., 1999. Teaching to Big Ideas. Solomon, M.Z. (Ed.), The Diagnostic Teacher: Constructing New Approaches to Professional Development. New York: Teacher College Press.

Shaanxi statistical bureau., 2014. Shaanxi Statistical Yearbook. China Statistics Press, Beijing.

Shi, Y., Zhang, L., Ma, Y., Yi, H., Liu, C., Johnson, N., Rozelle, S., 2015. Dropping Out of Rural China's Secondary Schools: A Mixed-methods Analysis. The China Quarterly, 224, 1048-1069.

Showers, B., Joyce, B., Bennett, B., 1987. Synthesis of Research on Staff Development: A Framework for Future Study and A State-of the-art Analysis. Educational Leadership, 45(3), 77–87.

Subirats, J., Nogales, I., 1989. Teachers, Schools, Educative Crisis Conditions of Teachers' Work in Bolivia. Santiago, Chile: OREALC/ UNESCO. (In Spanish)

Tatto, T., 1997. Limits and Constraints to Effective Teacher Education. International Handbook of Education and Development: Preparing Schools, Students and Nations for the Twenty-first Century, 213-230.

UNESCO Institute for Statistics. 2006. Teachers and Educational Quality: Monitoring Global Needs for 2015. International Institute for Educational Planning, UNESCO, Montreal.

Vegas, E., 2007. Teacher Labor Markets in Developing Countries. The Future of Children, 17(1), 219-232.

Villegas-Reimers, E., 2003. Teacher Professional Development: An International Review of the Literature. Paris: International Institute for Educational Planning.

Wang, X., Liu, C., Zhang, L., Luo, R., Shi, Y., Rozelle, S., and Sharbono, B., 2009. What is Keeping the Poor out of College? Enrollment Rates, Educational Barriers and College Matriculation in China. China Agricultural Economic Review, 3, 131-149.

Wang, H., Yang, C., He, F., Shi, Y., Qu, Q., Rozelle, S., and Chu, J., 2015. Mental health and dropout behavior: A cross-sectional study of junior high students in northwest rural China. International Journal of Educational Development, 41, 1-12.

World Bank. 2001. China: Challenges of Secondary Education. Second Education Series 22856, World Bank, Washington D.C.

Yan H., Wei, F., and Li, B., 2013. The Situation and Development of a Roadmap of Teacher Training. Teacher Education, 5(6), 43-50.

Ye, J., and Pang, L., 2011. Differentiated Childhoods: Impacts of Rural Labor Migration on Left-behind Children in China. The Journal of peasant studies, 38(2), 355-377.

Ye, J., Wang, Y., Zhang, K., Lu, J., 2006. Life affection of parents going out for work to left behind children. Chinese Rural Econ, 1, 57–65 (in Chinese).

Yi, H., Zhang, L., Luo, R., Shi, Y., Mo, D., Chen, X et al., 2012. Dropping Out: Why Are Students Leaving Junior High in China's Poor Rural Areas? International Journal of Educational Development, 32(4), 555-563.

Yoshikawa, H., Leyva, D., Snow, C. E., Treviño, E., Barata, M., Weiland, C., and Arbour, M. C., 2015. Experimental Impacts of a Teacher Professional Development Program in Chile on Preschool Classroom Quality and Child Outcomes. Developmental psychology, 51(3), 309.

Yoon, S., Duncan, T., Lee, Y., Scarloss, B., and Shapley, L., 2007. Reviewing the Evidence on How Teacher Professional Development Affects Student Achievement. Issues & Answers. REL 2007-No. 033. Regional Educational Laboratory Southwest (NJ1).

Zhou, C., Sylvia, S., Zhang, L., Luo, R., Yi, H., Liu, C., Shi, Y., Loyalka, P., Chu, J., Medina, A., Rozelle, S., 2015. China’s left-behind children: impact of parental migration on health, nutrition, and educational outcomes. Health Affairs, 34(11), 1964-1971.

Zuo E., Su, Q., 2012. The Investigation of Rural Middle School Core Physical Education Teachers Participating in Hunan Province “National Training Program”. *Nenjiang Technology,* 3(2), 20-25 (in Chinese).

Table 1: Comparisons of attrition between the treatment and control groups.

Dependent Variable: Attrition (1=student attrited, 0=student remained)	(1)	(2)
1. Treatment students (yes = 1; no = 0)	-0.02	-0.01
	(0.01)	(0.01)
School-pair fixed effect	Yes	Yes
Student Characteristics	NO	Yes
Observations	3289	3289
R-squared	0.001	0.004

Source: Authors' survey

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

Robust standard errors in parentheses

Table 2. Student and teacher characteristics at baseline

Variable	Treatment group		Control group		Difference: (1) - (2)		P-value
	(1)		(2)		(3)		(4)
	Mean	Sd	Mean	Sd			
Student Characteristics							
1. Standardized math test score (SD)	-0.00	[1.00]	0.00	[1.00]	-0.01	(0.06)	0.91
4. Student age (year)	9.67	[1.41]	9.59	[1.39]	0.09	(0.08)	0.28
5. Female student (1=yes, 0=no)	0.48	[0.50]	0.47	[0.50]	0.01	(0.02)	0.63
2. Student is a left-behind child (1=yes, 0=no)	0.15	[0.36]	0.15	[0.36]	-0.00	(0.01)	0.94
3. Student lives at school (1=yes, 0=no)	0.07	[0.25]	0.08	[0.27]	-0.01	(0.02)	0.47
6. Mother completed junior high school (1=yes, 0=no)	0.41	[0.49]	0.41	[0.49]	0.00	(0.02)	0.92
7. Father completed junior high school (1=yes, 0=no)	0.41	[0.49]	0.43	[0.50]	-0.02	(0.02)	0.34
Number of observations	1144		1922		3066		
Teacher Characteristics							
8. Teacher math teaching knowledge test score (SD)	-0.03	[1.16]	-0.11	[1.07]	0.08	(0.26)	0.77
9. Female teacher (1=yes, 0=no)	0.85	[0.36]	0.82	[0.39]	0.03	(0.07)	0.62
10. Teacher completed university (1=yes, 0=no)	0.18	[0.39]	0.30	[0.46]	-0.12	(0.08)	0.13
11. Teacher majored in math (1=yes, 0=no)	0.15	[0.36]	0.32	[0.47]	-0.17*	(1.00)	0.08
12. Teacher has highest rank (1=yes,0=no)	0.15	[0.36]	0.24	[0.43]	-0.09	(0.09)	0.30
13. Teaching experience (1=over 15 years, 0=equal or below 15 years)	0.21	[0.41]	0.46	[0.50]	-0.25**	(0.09)	0.01
Number of observations	34		50		84		

Source: Authors' survey

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

Table 3. The effect of the NTTP on student academic achievement (Shaanxi Pro
China)

Dependent variable: Endline student math score (SD)	(1)	(2)
	Unadjusted	Adjusted
1.Teacher participates in teacher training (1=yes, 0=no)	0.00	-0.07**
	(0.05)	(0.05)
2. Baseline student math score (SD)		0.52***
		(0.02)
Student characteristics [a]	NO	YES
Teacher characteristics [b]	NO	YES
School-pair fixed effect [c]	YES	YES
Constant	-0.00	1.61***
	(0.02)	(0.24)
Observations	3,066	3,066
R-squared	0.148	0.413

Source: Authors' survey

Notes:

a. The unadjusted model does not control for student characteristics or teacher characteristics, adjusted model does. The student characteristics include: student age, gender, whether the left-behind child, whether the student lives at school, whether the student's mother comple high school, and whether the student's father completed junior high school.

b. The teacher characteristics include: teacher gender, whether the teacher completed universi the teacher majored in math, whether the teacher has achieved the highest teaching rank, a experience.

c. The unadjusted model and the adjusted model both include a school-pair fixed effect. Since 34 pairs of treatment schools and control schools in our sample, we include a matrix made school pair dummy variables.

d. Additionally, we ran another analysis that considers the two types of control teachers by ad dummy variable that equals 1 if the teacher was a within-school control teacher and 0 othe results of this approach are substantively identical to the above estimates. They are reporte supplemental appendix available online (https://reap.fsi.stanford.edu/publication/impact-teacher-professional-development-prograr hievement-rural-china).

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

Cluster-robust standard errors adjusted for clustering at the school-pair level in parentheses.

Table 4. Heterogeneous effects of NTTP on student academic achievement (Shaanxi Province, China)

Dependent variable: Endline student math score (SD)	(1)	(2)	(3)	(4)
1.Teacher participates in teacher training (1=yes, 0=no)	-0.07**	0.18	-0.03	-0.08**
	(0.03)	(0.22)	(0.04)	(0.03)
2. Teacher participates in teacher training *Baseline student math test score	0.00			
	(0.03)			
3. Teacher participates in teacher training *Student age		-0.03		
		(0.02)		
4. Teacher participates in teacher training *Female student			-0.07	
			(0.06)	
5. Teacher participates in teacher training *Student is a left-behind child				0.10
				(0.08)
6. Baseline student math score (SD)	0.52***	0.52***	0.52***	0.52***
	(0.02)	(0.02)	(0.02)	(0.02)
7. Student age (years)	-0.17***	-0.15***	-0.16***	-0.16***
	(0.02)	(0.03)	(0.02)	(0.02)
8. Female student (1=yes, 0=no)	-0.04	-0.04	-0.01	-0.04
	(0.03)	(0.03)	(0.04)	(0.03)
9. Student is a left-behind child (1=yes, 0=no)	-0.01	-0.01	-0.01	-0.04
	(0.03)	(0.04)	(0.04)	(0.05)
Student characteristics [a]	YES	YES	YES	YES
Teacher characteristics [b]	YES	YES	YES	YES
School-pair fixed effect [c]	YES	YES	YES	YES
Constant	1.61***	1.53***	1.61***	1.62***
	(0.24)	(0.25)	(0.24)	(0.24)
Observations	3,066	3,066	3,066	3,066
R-squared	0.413	0.413	0.413	0.413

Source: Authors' survey. *** p<0.01, ** p<0.05, * p<0.1

Notes: a. The student characteristics include: whether the student lives at school, whether the student's mother completed junior high school, and whether the student's father completed junior high school.

b. The teacher characteristics include: teacher gender, whether teacher completed university, whether teacher majored in math, whether teacher achieved the highest teaching rank, and teaching experience.

c. The heterogeneous effects analyses include a school-pair fixed effect. Since there were 34 pairs of treatment schools and control schools in our sample, we include a matrix made up of 34 school pair dummy variables.

Cluster-robust standard errors adjusted for clustering at the school-pair level in parentheses.

Table 5. The effects of the NTTP on the math teaching knowledge of teachers

Dependent variable: Endline teacher math teaching knowledge test score (SD)	(1)	(2)
	Unadjusted	Adjusted
1. Teacher participates in teacher training (1=yes, 0=no)	0.55**	0.42**
	(0.24)	(0.20)
2. Baseline teacher math teaching knowledge test score (SD)		0.57***
		(0.12)
Teacher Characteristics [a]	NO	YES
School-pair fixed effect [b]	YES	YES
Constant	-0.27*	-0.15
	(0.16)	(0.44)
Observations [c]	82	82
R-squared	0.465	0.700

Source: Authors' survey

Notes:

a. The unadjusted model does not control for teacher characteristics while the adjusted model does. The teacher characteristics include: teacher gender, whether the teacher completed university, whether the teacher majored in math, whether the teacher has achieved the highest teaching rank, and teaching experience.

b. The unadjusted model and adjusted model both include a school-pair fixed effect. Since there were 34 pairs of treatment schools and control schools in our sample, we include a matrix made up of 34 school pair dummy variables.

c. Two teachers (one that participated in the NTTP and one that did not participate in the NTTP) were unable to complete the teacher math teaching knowledge test in the endline survey. Thus we have 82 (instead of the 84 in our sample) teachers in this analysis.

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

Cluster-robust standard errors adjusted for clustering at the school-pair level in parentheses.

Table 6. The effect of the NTTP on the teaching practices of teachers

Dependent variable: Endline teaching practice index (SD) (Report from Students)	(1) Unadjusted	(2) Adjusted
1. Teacher participates in teacher training (1=yes, 0=no)	0.03	-0.00
	(0.02)	(0.02)
2. Baseline teacher teaching practice index (SD)		0.32***
		(0.04)
Teacher Characteristics [a]	NO	YES
School-pair fixed effect [b]	YES	YES
Constant	-0.01	-0.07
	(0.01)	(0.04)
Observations	3066	3066
R-squared	0.006	0.170

Source: Authors' survey

Notes:

a. The unadjusted model does not control for teacher characteristics while the adjusted model does. The teacher characteristics include: teacher gender, whether the teacher completed university, whether the teacher majored in math, whether the teacher has achieved the highest teaching rank, and teaching experience.

b. The unadjusted model and adjusted model both include a school-pair fixed effect. Since there were 34 pairs of treatment schools and control schools in our sample, we include a matrix made up of 34 school pair dummy variables.

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

Cluster-robust standard errors adjusted for clustering at the school-pair level in parentheses.

Figure 1. Framework for how the NTTP affects student achievement