July 2017

The Effect of Teacher Looping on Student Achievement: Evidence from Rural China

Lei Wang, Mengying Wu, Yaojiang Shi, Yazhuo Chen, Prashant Loyalka, James Chu, Kaleigh Kenny, and Scott Rozelle

Abstract

One form of classroom organization that may influence student achievement is teacher looping – a classroom arrangement in which teachers stay with the same group of students for two or more consecutive years. We employ a student fixed-effects model to estimate the impact of teacher looping on student achievement in rural China. We find that students in looping classrooms perform significantly better than their peers in non-looping classrooms. We also find that poor and boarding students, in particular, benefit from looping classroom organization. The gender of the looping teacher does not appear to impact student outcomes.

Keywords: Rural China; Teacher Looping; Student Achievement; Boarding Students

---

Working Paper 318

July 2017

reap.fsi.stanford.edu

REAP Rural Education Action Program

1. Introduction

Policymakers and researchers are concerned with how the organization of schools and classrooms shapes student outcomes. A number of studies indicate that disparities in student outcomes may be due to aspects of school and classroom organization such as principal and teacher incentives (Darling-Hammond, 2000; Eberts et al., 2002; Branch, Hanushek, & Rivkin, 2013; O'Donnell and White, 2005), class size (Bosworth, 2014; Hoxby, 2000), and student sorting (Cohen-Zada, Gradstein, & Reuven, 2013; Koedel & Betts, 2011; Rothstein, 2009). Indeed, some researchers argue that adjusting the organization of schools and classrooms may be much more effective at improving student outcomes than merely increasing school resources per student (Hanushek, 2004; Jacob & Ludwig, 2009). Research also shows that the gap between the effects of different teacher assignments on individual student outcomes can be dramatic. For example, students who are assigned to ineffective teachers across several years have significantly lower levels of achievement and gains in achievement than those who are assigned to highly effective teachers over the same time period (Hanushek, Rivkin, & Kain, 2005; Chetty, Friedman, & Rockoff, 2014).

One aspect of the allocation of teaching resources that is important, but has seldom been investigated in empirical studies, is teacher looping – a specific classroom arrangement in which teachers stay with the same group of students for two or more consecutive years (Grant, Johnson, & Richardson, 1996). Recognizing the potential of teacher looping as a vehicle to better educate children from low socioeconomic environments (Oberman, 1997), both developed and developing countries have adopted teacher looping in the hopes of improving student achievement (Grant, Johnson, & Richardson, 1996). For example, looping has been widely implemented in European and Asian countries as diverse as Germany, New Zealand, Finland, Netherlands, Denmark, Sweden, Japan, China, Russia, and Israel (Bempechat et al., 2011; Cistone & Shneyderman, 2004; Crosser, 2005).

While looping is a relatively widespread practice, the empirical literature is mixed on its relationship to student achievement. Teacher looping is thought to be positively associated with student achievement for several reasons. First, teacher looping allows teachers to better understand students and build close relationships with students and their parents (Caauwe, 2009; Cistone & Shneyderman, 2004; Gaustad, 1998; Kane & Staiger, 2008; Nichols & Nichols, 2002; Little & Dacus, 1999; Rasmussen, 1998). Researchers have also demonstrated that looping teachers better understand the strengths, weaknesses, and personalities of their students than non-looping teachers (Zahorik & Dichanz, 1994). In addition, other research has shown that looping positively affects the attitudes of parents toward the educational environment of their children (Nichols & Nichols, 2002).

In contrast, the literature on looping also indicates some potential disadvantages. The greatest concern is that students may spend two or more years with an ineffective teacher in looping classrooms (Gaustad, 1998; Hitz, Somers, & Jenlink, 2007). This situation is troubling because teacher effectiveness is a strong determinant of student achievement (Darling-Hammond, 2016), and, therefore, is the key mechanism moderating the impact of teacher looping. Researchers also worry that looping teachers may have difficulty mastering curricula for multiple grades (Richards, 2010). There is evidence that a lack of specialization in teaching may be negatively associated with student achievement, because teachers need to spend additional time reacquainting themselves with lesson content and developing curricula each school year (Cistone & Shneyderman, 2004).

The current literature base on teacher looping primarily focuses on the association between its advantages/disadvantages and student outcomes, and few studies have sought to identify a causal link between looping and student achievement. For example, although some studies highlight the potential benefits teacher looping provides to student achievement, they have only used methods such as interviews and descriptive statistics (Burke, 1996; Chapman, 1999; Checkley, 1995; Forsten et al., 1997; Schieffer & Busse, 2001). In addition, a small number of studies have used multivariate regressions to show a positive correlation between looping and student achievement (Cistone & Shneyderman, 2004; Franz et al., 2010; Judge & Phillips, 2006; Nichols & Nichols,1999; Nichols & Nichols, 2002; Rodriguez & Arenz, 2007). Although multivariate analysis can establish a correlation between these two outcomes, most of these studies have not taken into consideration the unobserved heterogeneity between different teachers and different schools. Because of this, the results of these studies may be vulnerable to bias.

Given this evidence gap in the literature, the overall goal of this study is to examine whether there is a causal relationship between teacher looping and student achievement. To achieve this goal, we have three specific objectives. First, we describe the prevalence of teacher looping in rural China. Second, we seek to identify the causal relationship between teacher looping and student achievement. Last, we investigate whether there are heterogeneous impacts of teacher looping on the academic achievement of students. From the results of our heterogeneous analysis, we hope to be able to distinguish which groups benefit from, or are harmed by, teacher looping.

To achieve these goals, we analyze data on 3,784 elementary school students from 70 public elementary schools in rural northwest China.[1] We conduct this analysis using a cross-subject student fixed-effects model that controls for a large set of student and teacher covariates, allowing us to better control for the unobserved heterogeneity that may impact the results found in previous studies. In other words, this model removes potential sources of bias that can arise when there non-random sorting of students to looping and non-looping teachers and schools.

The rest of the paper is organized as follows. Section 2 provides background on educational outcomes and teacher looping in rural China. Section 3 describes our sample selection, data collection process, and analytical strategy. Section 4 presents the results on the impacts of teacher looping on student achievement and explores the potential mechanisms through which teacher looping affects student achievement. Section 5 discusses the results and section 6 concludes.

2. Rural Primary Education in China

The Chinese government recognizes that primary schooling is an important part of the education system in rural China. Because of this belief, the government eliminated tuition and fees for rural primary schools in 2007 in order to increase enrollment. It appears that this policy has

[1] The data were collected by researchers at the Rural Education Action Program at Stanford University between 2011 and 2012.

been successful, as the enrollment rate for primary school-age children was nearly 100% in 2015 (China National Bureau of Statistics, 2015).

Although recent measures have successfully increased rural primary school enrollment, similar progress has not been made in terms of academic achievement. For example, China's State Council developed a School Merger Program in 2000 in the hopes of improving the resources available to teachers and schools and, ultimately, the quality of education provided in rural primary schools. However, no research has found a measurable impact of this merger program on the academic performance of students (Liu et al., 2010). The need for improved educational outcomes in rural China is necessary, as it has been found that the academic performance of primary school students in rural areas fall far behind their urban peers (Song, 2010). Without more equitable human capital development across all of China, it is unlikely that the country will be able to sustain the growth rates that have driven economic development in recent decades.

Due to the persistence of this educational achievement gap between rural and urban students, researchers have sought to identify factors that contribute to the poor educational outcomes of rural students. Factors that have been identified as deficient in rural areas and related to education outcomes include nutrition, health, and vision care (Luo et al., 2011, 2012; Ma et al., 2014). One educational input that researchers could benefit from further examining is teacher quality, which has been identified as one of the most important inputs impacting student achievement in the international literature (Darling-Hammond, 2000). Research has also shown that teacher quality is important for improving student outcomes in China. For example, research has shown that teachers with the highest rank, with higher levels of educational attainment, and who receive incentives are more likely to have better educational outcomes (Chu et al., 2015; Zhang et al., 2009; Loyalka et al., 2015).

Despite the increased attention on teacher quality in rural China, virtually no research has examined the impacts of school organizational practices, such as teacher looping, on the academic achievement of students in rural China. This may be because teacher looping is not implemented in a systematic manner in China. How schools allocate or organize teachers is not decided by China's Ministry of Education, but rather through a decentralized decision-making process that is often carried out at the school level. Because of this decentralized management, there is considerable heterogeneity in the organizational practices of rural Chinese schools. For example, teacher looping is practiced in most primary and secondary schools, although the assignment of teachers to one grade level is also common (Liu, 1997; Mao, 1988; Wu & Chen, 2002). Due to evidence that teaching organizational practices are tied to student achievement, it is in the interest of the Chinese government to identify and systematize the school organizational practice that can lead to the best educational outcomes.

3. Research Design

3.1 Sampling

The panel data for this study are derived from a large-scale, longitudinal survey of 2,149 students and 218 teachers in Shaanxi Province (a northwestern province in China). All surveying was conducted by researchers associated with the Rural Education Action Program at Stanford University and the Center for Experimental Economics in Education at Shaanxi Normal University during baseline surveying in February 2011 and endline surveying in June 2012. Additionally, we use a panel of 3,784 students and 400 teachers constructed from midline (June 2011) and endline survey data to conduct robustness checks. Given the large sample size, this dataset can be considered broadly representative of the rural population in Shaanxi province. The data provide detailed information on students, households, teachers, and schools that were collected from student and teacher survey responses.

To select our sample of students, we first restricted our sampling frame to four randomly selected counties in one prefecture of Shaanxi Province. We decided to limit our sampling frame to this prefecture to ensure we are examined the effects of teacher looping in poor, rural areas and to further diminish any unobserved heterogeneity among students in our sample. Our sampled prefecture ranked 9th out of the 10 prefectures in Shaanxi Province in terms of GDP per capita in 2012 (China National Bureau of Statistics, 2013) and the three of the four sampled counties are nationally-designated poverty counties.

After selecting our sample counties, we obtained a comprehensive list of all elementary schools that had classes in grades one through six from the Department of Education for each county. We only selected schools of this type with larger student populations because policymakers informed us other schools may have been risk of being closed or merged during the school year. From this list, we selected 48 sample schools at the time of the baseline survey that the criteria for a large, interventional survey (not discussed in this paper). At the time of the midline survey (which was conducted just four months after the baseline), an additional 14 schools (for a total of 62 schools) were selected to increase the statistical power of the findings of the interventional study.

Finally, within the sample schools, we included all students in grades three and five in our sample at the time of the baseline survey, and extended our sample to include all students in grades two and four at the time of our midline survey. We conducted a follow-up survey one year after the midline survey (June 2012) with the same set of students when they were in grades three through six.

At the time of the endline survey, we also identified and surveyed all mathematics and Chinese teachers for grades three through six in our sample schools. The teacher questionnaire was distributed in June 2012, and included teacher characteristics and school characteristics for the school years 2011-2012. Altogether, we surveyed 2,149 students (clustered in 75 classes) and 218 of their teachers in baseline survey and 3,784 students (clustered in 130 classes) and 400 of their teachers in the midline survey.

3.2 Data Collection

In each wave, we conducted our survey in three blocks. In the first survey block, we collected information on the characteristics of sample students. To do so, we surveyed students to collect information such as student gender and boarding status (whether they board in school or not). We also obtained a measure of household socioeconomic status based on household ownership of a number of assets to construct a poverty index to distinguish between (economically) poor and non-poor students[2].

In the second block of the survey, we administered standardized tests in mathematics and Chinese language to all sample students. Both tests were carefully designed with assistance from educators in the local education bureau to ensure coherence with the national curriculum. We pre-tested the exam multiple times to ensure its relevance and that time limits were appropriate. Students were given 30 minutes to finish each examination. When we administered the exam in the sample schools, it was timed carefully and closely proctored by trained enumerators. For the purpose of our analysis, the mathematics and Chinese exam scores were normalized into z-scores by subtracting the mean and dividing by the standard deviation (SD) of the exam score distribution among the entire sample (for mathematics and Chinese separately).

In the third block of the survey we collected information on teacher characteristics and looping. Specifically, we asked questions of teachers about their gender, highest teaching ranking (whether a teacher achieved senior primary level or not), teaching awards (whether a teacher received a teaching award or not), college degree (whether a teacher attended college or not), number of hours spent teaching a specific subject (math or Chinese) every week, certificate completion (whether a teacher gets teaching certificate or not), and whether they hold a permanent or a temporary teaching position. We also asked questions of teachers to determine whether their

[2] We created the poverty index using polychoric principal components analysis (Kolenikov and Angeles, 2009).

educational background is appropriate for the academic subject they teach. To do so, we matched the college major of teachers with the subject they teach to identify whether they match or not. Finally, we asked teachers to assess their own Chinese language and math skills on a five-point Likert scale where “1” represents the lowest and “5” represents the highest knowledge levels. These self-reported skill levels serve as our measure of teaching ability.

In addition to collecting information on teacher characteristics, we also gathered information on teacher looping. To obtain information about teacher looping, we surveyed all mathematics and Chinese teachers that taught our sample students between February 2011 and June 2012. We then created a dummy variable equal to 1 if students were assigned to the same teacher during the two waves of data used for the analysis in question, and 0 if a student was assigned to different teachers. Importantly, no students in our sample changed classes during any semester when we conducted surveying.

3.3 Analytic Strategy

A main challenge of any quasi-experimental design for estimating the causal impact of teacher looping on student academic achievement is the potential selection bias due to the fact that students are not randomly assigned to their teachers. To provide an unbiased estimation of the effect of teacher dimensions on student academic achievement, researchers have used cross-subject student fixed effect models (Chu et al., 2015; Clotfelter et al., 2010). This is because a student fixed effect removes the potential impact of unobserved, subject-invariant characteristics on our outcomes measure. In other words, if a student is in a looping classroom for math and not for

Chinese, the difference between his/her achievement in math and Chinese as compared to that difference in performance of a non-looping student can be considered a less biased estimator than that found by merely comparing the relative performance of looping and non-looping students.

Following the approach of Clotfelter et al. (2010) and Chu et al. (2015), we use within-student variation across math and Chinese performance to identify causal impacts. To illustrate how the cross-subject student fixed effects model removes the potential effects of unobservable student characteristics on our outcome measure, we first examine the relationship between student achievement and teacher looping using a standard linear regression model:

$$\mathrm{Achievement}_{\mathrm{is,T}} = \alpha + \gamma \mathrm{Achievement}_{\mathrm{is,T-1}} + \beta \mathrm{Looping}_{\mathrm{is}} + \delta \mathrm{C}_{\mathrm{is}} + \lambda_{\mathrm{i}} + \varepsilon_{\mathrm{is}} \quad (1)$$

where student *i* takes classes Math or Chinese subjects *s*. $Achievement_{is,T}$ is the endline math or Chinese test of student *i* in subject *s* at time *T*, while $Achievement_{is,T-1}$ is baseline achievement of student *i* in subject *s* at time *T-1*. $Looping_{is}$ is a treatment variable indicating whether the student is in a looping classroom (1=yes); C_{is} is a vector of teacher characteristics that serve as controls. In this study, C_{is} is a vector representing teacher covariates, such as gender and class hours (see the full list of teacher covariates in Table 1). Individual student characteristics are not explicitly denoted in this model because we include λ_i, a student fixed effect, which controls for all time-invariant student characteristics. Finally, the symbol ε_{is} represents an error term that encompasses student, teacher, and classroom characteristics that vary across subjects. The other terms (α, γ, β, and δ) in the equation are coefficients (or vectors of coefficients) to be estimated.

We are most interested in β, which identifies the relationship between teacher looping and student achievement.

Because the student fixed effects λ_i are equivalent across both subjects, differencing Equation (1) for the two subjects yields an equivalent Equation (2) as follows:

$$\mathrm{Achievement}_{i1,T} - \mathrm{Achievement}_{i2,T} = \gamma\left(\mathrm{Achievement}_{i1,T-1} - \mathrm{Achievement}_{i2,T-1}\right) + \beta(\mathrm{Looping}_{i1} - \mathrm{Looping}_{i2}) + \delta(\mathrm{C}_{i1} - \mathrm{C}_{i2}) + (\varepsilon_{i1} - \varepsilon_{i2}) \qquad (2)$$

Unobserved teacher and classroom characteristics that vary across subjects are captured in the differenced error term $(\varepsilon_{i1} - \varepsilon_{i2})$. To obtain unbiased estimates of β, this model relies on the assumption that the error term $(\varepsilon_{i1} - \varepsilon_{i2})$ in Equation (2) must be uncorrelated with the treatment $(Looping_{i1} - Looping_{i2})$ and student achievement $(Achievement_{i1,T} - Achievement_{i2,T})$. We include $\left(Achievement_{i1,T-1} - Achievement_{i2,T-1}\right)$ and the vector $(C_{i1} - C_{i2})$ to account for this possibility that the error term and the treatment are in fact correlated. To test the consistency of our results, we run Equation (2) twice. The second time, we use our measure of the academic achievement of students collected at the time of the mid-line survey, rather than at the time of the baseline survey.

To conduct our heterogeneous analysis, we interact the treatment variable in Equation (2) with the student and teacher characteristics to examine whether there are heterogeneous impacts of teacher looping on student achievement. The variables we take into consideration in our heterogeneous analysis are student gender, boarding status, poverty status, and baseline test scores, as well as teacher gender.

The characteristics of math and Chinese language teachers are detailed in Table 1. From these data, we can see that math teachers are more likely to have majored in math than Chinese language teachers were to have majored in Chinese language (significant at the 1% level). In addition, math (as a subject) tends to be taught fewer times per week than Chinese (significant at the 10% level). We also find that Chinese language teachers are more likely to be female than math teachers (significant at the 10% level). Although we only find small differences between the characteristics of our math and Chinese language teachers, we control for all these observable teacher characteristics in our analysis.

4. Results

4.1 Prevalence of teacher looping in rural China

As can be seen in Table 2, looping teachers are more common in our sample than non-looping teachers (rows 1 and 2). Specifically, among the 400 teachers in our sample, 289 of them are looping teachers (72% of the sample) and 111 of them are non-looping teachers (28% of the sample). To be clear, in the context of our study (given our data), this means that in the case of 72% of our teachers, the teachers taught the same students in either math or Chinese language in both years of our sample. The result show that the practice of teacher looping is a common organizational practice in rural Chinese schools and that, in our sample sites, it is by far the most common teaching organizational practice.

We also find that Chinese language teachers are more likely than math teachers to be looping teachers. In total, there are 192 Chinese language teachers and 208 math teachers in our sample (Table 2, rows 3 to 6). Within the subsample of Chinese language teachers, 151 are looping teachers (or 79% of Chinese language teachers) and 41are non-looping teachers (21% of Chinese language teachers). For math teachers in our sample, 138 are looping teachers (66% of math teachers) and 70 are non-looping teachers (34% of math teachers).

Table 2 also shows that, although there are differences in the prevalence of looping and non-looping teachers, there are few significant differences in observable teacher characteristics between these two groups (rows 7, 10, 12-16). For example, we find a small, insignificant difference in the share of teachers with the highest rank (28% of looping teachers vs. 27% of non-looping teachers) and we find no difference in the share of teachers who are certificated between looping and non-looping teachers (both 98%). Similar results also are found for the average weekly hours looping teachers and non-looping teachers spend working on and preparing lessons for each class they teach (8.84 hours and 8.85 hours a week, respectively).

We also found no significant differences in scores for the academic skills self-assessments between our looping and non-looping teacher groups. Specifically, the average Chinese language skill score for looping Chinese language teachers is 3.63 and 3.73 for non-looping Chinese language teachers, representing only a 0.10-point difference (Table 2, row 15). Likewise, the average math skill score for looping math teachers is 3.94 and 3.88 for non-looping math teachers, and this difference (0.06 points) is insignificant (Table 2, rows 16).

While the observable characteristics of looping and non-looping teachers are mostly balanced, there are a several variables for which there are significant differences. Specifically, we find a difference in terms of receiving a teaching award, holding a college degree, and gender (Table 2, rows 8, 9 and 11). Specifically, we find that 45% of looping teachers in our sample have received a teaching award, while only 35% of non-looping have received an award. This difference is significant at the 10% level. We also find that in terms of holding a college degree, 10 percentage points more non-looping teachers hold a degree (32% of non-looping teachers) than looping teachers (22% of looping teachers –significant at 5% level). In terms of gender, there are 17 percentage points more female looping teachers (59% of looping teachers) than there are female non-looping teachers (42% of non-looping teachers– significant at 1% level). Because of these somewhat conflicting results (that is, that looping teachers are more likely to have won awards, but less likely to have a college degree), it is not possible to say if the quality of one group of teachers is higher than the other.

4.2 The impact of teacher looping on student achievement

To evaluate the impact of teacher looping on student achievement, we present the results of our cross-subject student fixed-effect model in Table 3. In this table, we control for the academic scores of students at either the time of the baseline (Columns 1 and 2) or midline (Columns 3 and 4) surveys. After controlling for the scores of students at the time of the baseline survey, we can see that teacher looping has a positive and significant impact on the academic achievement of students. Specifically, teacher looping increases student achievement by 0.14 SD (significant at the 5%

level). After controlling for potential confounding variables, we find that the adjusted result of 0.19 SD is significant at the 10% level (Column 2).

After adjusting our model to control for student scores at the time of the midline survey, we find that the results presented in Columns 3 and 4 are essentially the same as those in Columns 1 and 2. The results in Column 3 show that being in a classroom with a looping teacher is associated with average test scores 0.20 SD higher than those of students in a classroom with a non-looping teacher (significant at the 1% level). After controlling for covariates, we find that the adjusted result shows that teacher looping is correlated with average test scores 0.17 SD higher than those of the comparison group (significant at the 1% level).

Taken together, the results shown by Table 3 suggest that students in looping classrooms perform significantly better than their peers in non-looping classrooms in rural primary schools in China. This is an important finding since, to our knowledge, this is the first study to establish a causal relationship between teacher looping and student performance. In this way, the results help validate findings of existing qualitative/descriptive studies that claim that looping is an effective teaching practice (Burke, 1996; Chapman, 1999; Checkley, 1995; Forsten et al., 1997; Schieffer & Busse, 2001). These results also undermine the concerns of some researchers that teaching looping will negatively impact instruction quality. Teaching quality could suffer if looping teachers lack specialized knowledge on the material they teach (as compared to a situation where they teach the same material every school year—Cistone & Shneyderman, 2004). In other words, it is thought that looping teachers might need to spend time reacquainting themselves with academic content and developing curricula each school year, which could undermine their teaching effectiveness. However, our findings do not support this idea.

4.3 Heterogeneity analysis

To further understand what types of students benefit from teacher looping organizational practices, we estimate the heterogeneous impacts of teacher looping on student achievement along different student and teacher characteristics. The results of our heterogeneous analysis are presented in Table 4.

In terms of student baseline achievement, we find that teacher looping does not impact the academic achievement of the lowest-performing students any more than it does for other students. Although previous research has suggested that more positive student-teacher relationships (a hypothesized result of teacher looping) have a more positive effect for low-performing students and students with other learning issues (Baker, 2006), we do not find this result using our data.

We also find no statistically significant differences along student gender. According to our results, although the point estimate of having a looping teacher reduces the point estimate of the academic achievement of male students (0.22-0.05=0.17 SD) relative to that of female students (0.22 SD) by 0.05 standard deviations (Table 4, column 2), these findings are not statistically significant. This result supports those of previous research that found teacher looping was not differentially effective for students of different genders (Bogart, 2002; Patterson, 2003).

In contrast, our results demonstrate that teaching looping has a positive and significant impact on the academic achievement of poorer students. We find that when students whose household assets value is below the median are placed in classrooms with looping teachers, their academic achievement improves by 0.17 standard deviations (relative to non-poor student achievement). Importantly, this result is significant at the 5% level (Table 4, column 3). This result is supported by a finding that low-income families reported more positive perceptions of looping classrooms than families of higher income (Nichols & Nichols, 2003). One reason this may be the case is that the looping classroom model offers students and parents from lower socioeconomic groups a longer amount of time to build close relationships with teachers. With looping teachers, these students and parents may be more comfortable communicating with teachers after interacting with them for an extended period of time. As can be seen in Appendix Table 1, this result holds when we examine the heterogeneous effects using midline rather than baseline scores.

Finally, the heterogeneous analysis also suggests that teaching looping has a more significant positive impact on the academic achievement of boarding students as compared to non-boarding students. Specifically, we find that the academic achievement of boarding students is 0.34 SD higher than that of non-boarding students in looping classrooms (significant at the 1% level—Table 4, column 4). In fact, it may not surprising to find that boarding students benefit more from being placed in looping classrooms than their peers. Because boarding students live at school during the week, they have more opportunities to interact with their teachers than non-boarding students. In addition, the relationship that boarding students have with their teachers is likely more important than it is for non-boarding students who are in the care of their parents or another caregiver during the week. When we examine the case using midline rather than baseline examination scores, we again find that boarding students benefit more from teacher looping than non-boarding students (Appendix Table 1).

In our final empirical exercise (Table 5), we seek to identify whether being placed in a looping classroom with a teacher of the same gender positively affects student academic performance. That is, we seek to identify if girls perform better in looping classrooms with female teachers, and if boys perform better in looping classrooms with male teachers. The literature base suggests that this could be the case, as it has been argued that assignment to a teacher of the same gender can improve student academic achievement (Carrell et al., 2010; Dee, 2007; Metzler & Woessmann, 2012). However, our results presented in Table 5 show that being placed in a looping classroom with a teacher of the same gender has no additional impact on student achievement.

5. Discussion

The call for developing a high-quality educational system across all of China continues to intensify as the country continues to develop and move towards a modern, knowledge-based economy. Given the educational deficiencies in rural areas, rural education reform has long been a policy concern in China. One school organizational practice that has been examined as a potential method to improve student achievement is teacher looping (Caauwe, 2009; Rodriguez & Arenz, 2007). While the potential effectiveness of teacher looping has been raised in the literature, no study, to our knowledge, has identified a causal relationship between looping and student learning outcomes.

In this study, we employed a cross-subject student fixed effects model to analyze the impact of teacher looping on student learning. We did this by comparing the relative math and Chinese language scores of individual students who just happen to have one teacher that is a looping teacher (either their math or Chinese language teacher) and whose other teacher is non-looping teacher. In this manner, we're able to isolate the effect of teacher looping on academic achievement while controlling for all time-invariant student characteristics. Using this strategy, we find a positive and significant effect of teacher looping on student academic achievement.

While we find that looping benefits educational outcomes, in general, the results of our heterogeneous analysis suggest that teacher looping has a more positive effect for some subgroups of children than it does for others. For example, our results showed that boarding students benefit more from teacher looping than do non-boarding students. This result is promising considering the current state of rural education in China. Specifically, the number of boarding students has increased rapidly since the implementation of rural primary school Merger Program and surpassed 30 million students by 2008 (Ministry of Education, 2008). If this trend continues (as it very well may do), teacher looping may be able to serve as an effective method to improve education quality in rural areas.

We also found that teacher looping more positively affects the educational outcomes of poorer students in our sample (those with household assets below the median value). For this reason, it may be especially beneficial for teacher looping to be implemented in the poorest regions of rural

China. For example, if schools in nationally designated poverty counties were encouraged or incentivized to implement teacher looping, this could offer educational benefits to their students.

Although we have found that teacher looping does generally benefit the education of rural Chinese students, further steps must be taken to ensure that looping is well implemented and teacher attrition is kept at a low level. It might be useful for the government to increase funding for rural education, which could translate to increased salaries and improved benefits that could encourage teachers to stay in their positions and with the students in their classrooms. In addition, policymakers and school administrators should encourage systematic lesson planning and provide professional development for looping teachers in order to ensure that this organizational practice is implemented effectively. It is likely that careful curriculum preparation and practical training are imperative to the effective multi-grade looping. Additionally, if teachers are able to share instruction resources amongst themselves, this could help reduce the burden becoming reacquainted with lesson content and devising lesson plans.

Although our research was capable of establishing a causal link between teacher looping and student achievement, improvements can still be made to this research. Future work should extend this research to schools in other Chinese provinces beyond Shaanxi. In addition, it may be worthwhile to examine the effects of teacher looping over a longer horizon to examine whether these positive findings hold over time.

Reference

Baker, J. A. (2006). Contributions of teacher–child relationships to positive school adjustment during elementary school. *Journal of School Psychology*, *44*(3), 211-229.

Bempechat, J., Mirny, A., Li, J., Wenk, K., & Holloway, S. (2011) Learning together: The educational experiences of adolescents in Moscow. In D. McInerney & R. Walker (Eds.), *Sociocultural theories of learning and motivation looking back, looking forward*. Hong Kong: Information Age.

Bogart, V. S. (2002). The Effects of Looping on the Academic Achievement of Elementary School Students. Electronic Theses and Dissertations Paper 707. School of Graduate Studies, East Tennessee State University. http://dc.etsu.edu/etd/707.

Bosworth, R. (2014). Class size, class composition, and the distribution of student achievement. *Education Economics, 22,* 141-165.

Branch, G., Hanushek, E., & Rivkin, S. (2013). School leaders matter: Measuring the impact of effective principals. *Education Next*, *13(*1), 62-69.

Burke, D. L. (1996). Multi-year teacher/student relationships are a long-overdue arrangement. *Phi Delta Kappa, 77*(5), 360-361.

Caauwe, C. M. (2009). The Impact of Looping Practices on Student Achievement at a Minnesota Inner City Elementary School: A Comparison Study. Doctoral Dissertation, Saint Mary's University of Minnesota. Retrieved from: http://files.eric.ed.gov/fulltext/ED512646.pdf (accessed 29 April 2017).

Carrell, S. E., Page, M. E., & West, J. E. (2010). Sex and science: How professor gender perpetuates the gender gap. *The Quarterly Journal of Economics*, *125*(3), 1101-1144.

Chapman, J. (1999). Same teacher, different years: Facts about looping. *Young Children, 54(3),* 80-83.

Checkley, K. (1995). Multiyear education: Reaping the benefits of "looping." *Education Update, 37*(8), 1-6.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *The American Economic Review*, *104*(9), 2633-2679.

China National Bureau of Statistics. (2013). China Statistical Yearbook. China Statistics Press, Beijing.

China National Bureau of Statistics. (2015). *China Statistical Yearbook.* Beijing: China Statistics Press.

Chu, J. H., Loyalka, P., Chu, J., Qu, Q., Shi, Y., & Li, G. (2015). The impact of teacher credentials on student achievement in China. *China Economic Review*, *36*, 14-24.

Cistone, P., & Shneyderman, A. (2004). Looping: An Empirical Evaluation. *International Journal of Educational Policy, Research, and Practice: Reconceptualizing Childhood Studies*, *5*(1), 47-61.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2010). Teacher Credentials and Student Achievement in High School: A Cross-Subject Analysis with Student Fixed Effects. *Journal of Human Resources*, *45*(3), 655-681.

Cohen-Zada, D., Gradstein, M., & Reuven, E. (2013). Allocation of students in public schools: Theory and new evidence. *Economics of Education Review*, *34*, 96-106.

Crosser, S. (2005). *What do we know about early childhood education? Research based practice*. Clifton Park, NY: Thomson Delmar Learning.

Darling-Hammond, L. (2000). Teacher Quality and Student Achievement: A Review of State Policy Evidence. *Education Policy Analysis Archives, 8(1)*, 1-44.

Darling-Hammond, L. (2016). Research on Teaching and Teacher Education and Its Influences on Policy and Practice. *Educational Researcher*, *45*, 83-91.

Dee, T. S. (2007). Teachers and the gender gaps in student achievement. *Journal of Human Resources*, *42*(3), 528-554.

Eberts, R., Hollenbeck, K., & Stone, J. (2002). Teacher Performance Incentives and Student Outcomes. *The Journal of Human Resources*, *37*(4), 913–927.

Forsten, C., Grant, J., Johnson, B., & Richardson, I. (1997). *Looping Q&A: 72 Practical Answers to Your Most Pressing Questions*. Crystal Springs Books, Peterborough, NH.

Gaustad, J. (1998). Implementing Looping. Eric Digest No. 123. Retreived from: http://files.eric.ed.gov/fulltext/ED429330.pdf (accessed 1 May 2017).

Franz, D. P., Thompson, N. L., Fuller, B., Hare, R. D., Miller, N. C., and Walker, J. (2010). Evaluating mathematics achievement of middle school students in a looping environment. *School Science and Mathematics, 110(6)*, 298-308.

Grant, J., Johnson, B., & Richardson, I. (1996). *The looping handbook: Teachers and students progressing together.* Peterborough, NH: Crystal Springs Books.

Hanushek, E. (2004). *Distributional Outcomes of the Organization of U.S. Schools: Peers, School Quality, and Achievement*, in E. Hanushek, Schooling and human capital formation in the global economy: Revisiting the equity-efficiency quandary.

Hanushek, E., S. Rivkin, & J. Kain. (2005). Teachers, schools, and academic achievement. *Econometrica, 73*(2), 417-458.

Hitz, M., Somers, M., & Jenlink, C. (2007). The looping classroom: Benefits for children, families, and teachers. *Young Children, 62,* 80-84.

Hoxby, Caroline M. (2000). The Effects of Class Size on Student Achievements: New Evidence from Population Variation. *Quarterly Journal of Economics, 115*(4), 1239-1285.

Jacob, B., and Ludwig, J. (2009). "Improving educational outcomes for poor children." In M. Cancian & S. Danziger (Eds.), *Changing Poverty, Changing Policies*. New York: Russell Sage Foundation Press.

Judge, S. & Phillips, M. (2006). Does Looping Make a Difference? The Impact of Preschool Looping on Child Outcomes. *NHSA Dialog, 9(1)*, 12-21.

Kane, T. J., & Staiger, D. O. (2008). *Estimating teacher impacts on student achievement: An experimental evaluation* (No. w14607). National Bureau of Economic Research.

Koedel, C. and Betts, J. R. (2011). Does Student Sorting Invalidate Value-Added Models of Teacher Effectiveness? An Extended Analysis of the Rothstein Critique. *Education Finance and Policy, 6*(1), 18-42.

Kolenikov, S., & Angeles, G. (2009). Socioeconomic Status Measurement with Discrete Proxy Variables: Is Principal Component Analysis a Reliable Answer? *Review of Income and Wealth*, 55(1), 128-165.

Little, T. S., & Dacus, N. B. (1999). Looping: Moving Up with the Class. *Educational Leadership*, *57*(1), 42-45.

Liu, C., Zhang, L., Luo, R., Rozelle, S., & Loyalka, P. (2010). The effect of primary school mergers on academic performance of students in rural China. *International Journal of Educational Development*, *30*(6), 570-585.

Loyalka, P., Sylvia, S., Liu, C., Chu, J., & Rozelle, S. (2015). Teaching to the Tails: Teacher Performance Pay and the Distribution of Student Achievement. *Society for Research on Educational Effectiveness*.

Luo, R., Zhang, L., Liu, C., Zhao, Q., Shi, Y., Miller, G., ... & Martorell, R. (2011). Anaemia among students of rural China's elementary schools: prevalence and correlates in Ningxia and Qinghai's poor counties. *Journal of Health, Population and Nutrition*, 471-485.

Luo, R., Shi, Y., Zhang, L., Liu, C., Rozelle, S., Sharbono, B., ... & Martorell, R. (2012). Nutrition and educational performance in rural China’s elementary schools: Results of a randomized control trial in Shaanxi Province. *Economic development and cultural change*, *60*(4), 735-772.

Ma, X., Zhou, Z., Yi, H., Pang, X., Shi, Y., Chen, Q., …& Congdon, N. (2014). Effect of providing free glasses on children’s educational outcomes in China: cluster randomized controlled trial. *BMJ, 349,* g5740.

Mao, R. (1988). The Disadvantage of Teacher Looping [in Chinese]. *Yunnan Education: Basic Education Edition* (Z2).

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 (2008). The great progress of compulsory education in rural China, 2008-2-26, from http://www.moe.edu.cn/edoas/website18/83/info1203993432703183.htm (accessed 12 Feb 2017).

Nichols, G. W., & Nichols, J. D. (1999). Looping: The impact on parental attitudes in the educational environment. *International Journal of Educational Reform, 8(3)*, 274-279.

Nichols, J. D., & Nichols, G. W. (2002). The impact of looping and non-looping classroom environments on parental attitudes. *Educational Research Quarterly, 26(1)*, 23.

Nichols, J. D., & Nichols, G. W. (2003). The impact of looping classroom environments on parental attitudes. *Preventing School Failure: Alternative Education for Children and Youth*, *47*(1), 18-25.

Oberman, I. (1997). The Mystery of Waldorf: A Turn-of-the-Century German Experiment on Today's American Soil. Paper presented at the American Education Research Association

Annual Meeting, Chicago, 1997. Stanford, CA: Center for Research on the Context of Teaching, Stanford University.

O'Donnell, R. J. and White, G. P. (2005). Within the Accountability Era: Principals' Instructional Leadership Behaviors and Student Achievement. *NASSP Bulletin, 89*, 56-71.

Patterson, G. W. (2003). A comparison of multi-year instructional programs (looping) and regular education program utilizing scale scores in reading. (Doctoral dissertation, University of Florida, 2003).

Rasmussen, K. (1998). Looping: Discovering the benefits of multiyear teaching. *Education Update*, *40*(2), 1-4.

Richards, V. (2010). Life in Classrooms. In C. Kridel (Eds.), *Encyclopedia of curriculum studies*. Thousand Oaks, CA: Sage Publications.

Rodriguez, C., & Arenz, B. (2007). The effects of looping on perceived values and academic achievement. *ERS Spectrum, 25(3)*, 43–55.

Rothstein, J. (2009). Student sorting and bias in value-added estimation: selection on observables and unobservables. *Education Finance and Policy, 4*(4), 537-571.

Wu, G., & Chen, Y. (2002). Comparison Between Regular Shift system and Semester Shift System Under the Condition of Quality Education [in Chinese]. *Journal of Dalian Education University*, 18(4):18-20.

Schieffer, J., & Busse, R.T. (2001). Low SES minority fourth-graders' achievement from an urban public Waldorf and comparison school. *Waldorf Education Research Institute Research Bulletin, 6*, 18-24.

Song, N. Q. (2010). Urban-Rural Difference in Math Achievement of 4^{th}-6^{th} Grade Chinese Children and Adolescents [in Chinese]. *National Conference on Psychology*.

Zahorik, J. & Dichanz, H. (1994). Teaching for understanding in German schools. *Educational Leadership, 5,* 75-77.

Zhang, H., Zhang, L., Luo, R., & Liu, C. (2009). The impact of primary educational resource on education outcomes in rural China [in Chinese]. *Journal of Education and Economy*, 4, 1-6.

Table 1. Average math and Chinese language teacher characteristics

	Math Teacher	Chinese Language Teacher	Difference	P-value
	(1)	(2)	(1) - (2)	
Teacher Characteristics				
Observations	115	103	--	--
Share of total	0.53	0.47	0.06	--
Highest ranking (y/n)	0.30	0.20	0.10	0.12
Has teaching award (y/n)	0.41	0.45	-0.04	0.57
College degree (y/n)	0.27	0.26	0.01	0.86
Class hours	8.51	9.35	-0.84	0.09
Female (y/n)	0.45	0.58	-0.13	0.05
Right major to teaching subject (y/n)	0.09	0.01	0.08	0.01
Certificated teachers (y/n)	0.97	0.97	0.00	0.81
Official teachers (y/n)	0.95	0.94	0.01	0.85

Table 2. Teacher characteristics between looping and non-looping group

	Looping	Non-looping	Difference	P-value
	(1)	(2)	(1) - (2)	
Prevalence of Looping				
Observations in sample	289	111	178	--
Share of all teachers	0.72	0.28	0.44	--
Number of Chinese language teacher	151	41	110	--
Share of Chinese language teacher	0.79	0.21	0.58	--
Number of math teacher	138	70	68	--
Share of math teacher	0.66	0.34	0.32	--
Teacher Characteristics				
Highest ranking (y/n)	0.28	0.27	0.01	0.90
Has teaching award (y/n)	0.45	0.35	0.10	0.07
College degree (y/n)	0.22	0.32	-0.10	0.04
Class hours	8.84	8.85	-0.01	1.00
Female (y/n)	0.59	0.42	0.17	0.00
Right major to teaching subject (y/n)	0.05	0.08	-0.03	0.21
Certificated teachers (y/n)	0.98	0.98	0.00	0.86
Official teachers (y/n)	0.95	0.95	0.00	0.93
Chinese language skill[a] (1-5)	3.63	3.73	-0.10	0.51
Math skill[b] (1-5)	3.94	3.87	0.07	0.48

[a] Chinese language teachers were asked to assess Chinese language skill by themselves.

[b] Math teachers were asked to assess math skill by themselves.

Table 3. The impact of teacher looping on student achievement (using student fixed effect model).

Dependent variable: Endline student achievement (standard deviations)	(1)	(2)	(3)	(4)
Looping (y/n)	0.14**	0.19*	0.20***	0.17***
	(0.07)	(0.10)	(0.06)	(0.06)
Baseline score (s.d.)	0.18***	0.16***		
	(0.03)	(0.03)		
Mid-line score (s.d.)			0.21***	0.20***
(End of spring 2011)			(0.02)	(0.02)
Right major (y/n)		0.03		0.00
		(0.07)		(0.09)
College degree (y/n)		-0.11		-0.09
		(0.09)		(0.07)
Receiving teaching award (y/n)		0.07		0.08
		(0.07)		(0.05)
Highest ranking (y/n)		-0.09		-0.04
		(0.08)		(0.06)
Teacher gender (1=female)		0.05		0.13**
		(0.11)		(0.05)
Official teacher (y/n)		0.06		-0.08
		(0.14)		(0.10)
Certificated teacher (y/n)		0.10*		0.15**
		(0.05)		(0.07)
Class hours		0.01		0.01
		(0.01)		(0.01)
Constant	-0.12**	-0.43*	-0.13***	-0.32**
	(0.05)	(0.23)	(0.04)	(0.13)
Student fixed effect	Yes	Yes	Yes	Yes
Observations	4298	4298	7568	7568

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in parentheses clustered at classroom level.

Column (1) and (2) use baseline score. The baseline score is the normalized score on the math/Chinese test that is given to grade3 and grade5 students at baseline.

Column (3) and (4) use mid-line score. The mid-line score is the normalized score on the math/Chinese test that is given to grade2, grade3, grade4, grade5 students at end of spring semester 2011.

Table 4. Impact of teacher looping on student achievement of third grade and fifth grade (including interactions with boarding student, student low performance, poor student and student gender) by using student fixed effect model.

Dependent variable: Endline student achievement (standard deviations)	(1)	(2)	(3)	(4)
Looping (y/n)	0.19*	0.22	0.11	0.01
	(0.10)	(0.16)	(0.10)	(0.10)
Baseline score (s.d.)	0.16***	0.16***	0.16***	0.16***
	(0.05)	(0.03)	(0.03)	(0.03)
Looping * Baseline score (s.d.)	0.01			
	(0.05)			
Looping *Student gender		-0.05		
(Student gender:1=male; 0=female)		(0.16)		
Looping * Poor			0.17**	
(Poor: 1=family wealth lower than the median; 0=otherwise)			(0.07)	
Looping * Boarding				0.34***
(boarding :1=student is boarding ;0=otherwise)				(0.12)
Constant	-0.43*	-0.43*	-0.43*	-0.40*
	(0.23)	(0.23)	(0.23)	(0.23)
Teacher covariates controlled	Yes	Yes	Yes	Yes
Student fixed effect	Yes	Yes	Yes	Yes
Observations	4298	4298	4298	4298
T-test		Female	Non-poor	Non-boarding
Treatment effect		0.22	0.11	0.01
P-value		0.34	0.47	0.92
T-test		Male	Poor	Boarding
Treatment effect		0.17	0.28*	0.36**
P-value		0.25	0.06	0.03

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in parentheses clustered at classroom level.

The baseline score is the normalized score on the math/Chinese test that is given to grade 3 and grade 5 students at baseline.

Table 5. Impact of teacher looping on female versus male student achievement (including interactions with teacher gender) by using student fixed effect model.

Dependent variable: Endline student achievement (standard deviations)	(1) Male student	(2) Female student	(3) Male student	(4) Female student
Looping (y/n)	0.08	0.16	0.15**	0.18
	(0.11)	(0.17)	(0.07)	(0.11)
Baseline score (s.d.)	0.12***	0.21***		
	(0.04)	(0.04)		
Mid-line score (s.d.)			0.21***	0.19***
			(0.02)	(0.03)
Looping *Teacher gender (teacher gender: 1=female; 0=male)	0.29 (0.25)	0.26 (0.31)	0.05 (0.12)	0.01 (0.13)
Teacher gender (1=female;0=male)	-0.23 (0.0.21)	-0.11 (0.29)	0.01 (0.10)	0.21* (0.11)
Constant	-0.36	-0.36	-0.31**	-0.33
	(0.23)	(0.29)	(0.13)	(0.22)
Teacher covariates controlled	Yes	Yes	Yes	Yes
Student fixed effects	Yes	Yes	Yes	Yes
Observations	2264	1908	4022	3400

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in parentheses clustered at classroom level.

Column (1) and (2) use baseline score. The baseline score is the normalized score on the math/Chinese test that is given to grade3 and grade5 students at baseline.

Column (3) and (4) use mid-line score. The mid-line score is the normalized score on the math/Chinese test that is given to grade2, grade3, grade4, grade5 students at end of spring semester 2011.

Appendix 1． Impact of teacher looping on student achievement of third grade to fifth grade (including interactions with boarding student, student low performance, poor student and student gender) by using student fixed effect model.

Dependent variable: Endline student achievement (standard deviations)	(1)	(2)	(3)	(4)
Looping (y/n)	0.18***	0.21**	0.16	0.09
	(0.06)	(0.09)	(0.06)	(0.07)
Mid-line Score (s.d.)	0.16***	0.20***	0.20***	0.20***
(End of Spring 2011)	(0.03)	(0.02)	(0.03)	(0.02)
Looping * Mid-line score (s.d.)	0.06			
	(0.04)			
Looping *Student gender		-0.07		
(Student gender:1=male; 0=female)		(0.09)		
Looping * Poor			0.03	
(Poor: 1=family wealth lower than the median; 0=otherwise)			(0.10)	
Looping * Boarding				0.18**
(boarding :1=student is boarding; 0=otherwise)				(0.09)
Constant	-0.32**	-0.32**	-0.32*	-032**
	(0.13)	(0.13)	(0.19)	(0.13)
Teacher covariates controlled	Yes	Yes	Yes	Yes
Student Fixed Effect	Yes	Yes	Yes	Yes
Observations	7568	7568	7568	7568
T-test		Female	Non-poor	Non-boarding
Treatment effect		0.21*	0.16	0.09
P-value		0.10	0.14	0.40
T-test		Male	Poor	Boarding
Treatment effect		0.14	0.19**	0.27***
P-value		0.11	0.04	0.01

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in parentheses clustered at classroom level.

The mid-line score is the normalized score on the math/Chinese test that is given to grade 2, grade 3, grade 4, grade 5 students at end of spring semester 2011.