I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11347

Student Feedback, Parent-Teacher Communication, and Academic Performance: Experimental Evidence from Rural China

W Stanley Siebert
Xiangdong Wei
Ho Lun Wong
Xiang Zhou

FEBRUARY 2018

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11347

Student Feedback, Parent-Teacher Communication, and Academic Performance: Experimental Evidence from Rural China

W Stanley Siebert
University of Birmingham and IZA

Xiangdong Wei
Lingnan University

Ho Lun Wong
Lingnan University

Xiang Zhou
University of Birmingham

FEBRUARY 2018

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Student Feedback, Parent-Teacher Communication, and Academic Performance: Experimental Evidence from Rural China*

This study reports a randomized controlled trial to improve teacher-student-parent feedback, conducted in a rural county in China with many left-behind children. Data are collected from over 4,000 primary schoolchildren (8 to 10 years old) over two school terms. We find that bi-weekly student feedbacks using our special scorecard of schoolwork and behavior improve mathematics results by 0.16 to 0.20 standard deviations, with 0.09 for language. Communicating these assessments also to parents produces further large mathematics benefits for young left-behind children, about 0.30 standard deviations. A low-cost investment in better feedback thus brings significant achievement gains especially for disadvantaged children.

JEL Classification: C93, I21, J24

Keywords: student assessment, parent-teacher communication, academic performance, randomized controlled trial, rural China

Corresponding author:
W Stanley Siebert
Birmingham Business School
University of Birmingham
Birmingham, B15 2TT
United Kingdom

E-mail: w.s.siebert@bham.ac.uk

* We are grateful for the support from the General Research Fund of Research Grant Council in Hong Kong (Ref. No. LU342013) and the ESRC (Ref. No. RES-000-22-3654). We are also grateful for comments made by participants at the 2017 meetings of the Econometric Society in Asia, and the Asian and Australasian Society of Labor Economists.

NON-TECHNICAL SUMMARY

How to improve the education of disadvantaged children is an important issue, and teacher-student feedback has been advanced as a potent method. This research experimentally evaluates a simple form of such feedback, whereby the teacher discusses the student's academic and classroom performance every 2 weeks. For a subset of the students, the feedback results are also communicated with the parents. The research takes place in rural Hunan, a low-income area of central China where a large proportion of the school children are "left-behind", that is, have both parents working far away in the cities. 4,000 primary school children are involved in grades 3 and 5 (i.e. 8–10 years old).

Care is needed with the feedback method, since students with negative feedback (performance less than target) can feel discouraged. Discouragement is particularly likely for the left-behind children who lack parental support e.g. with homework. At the same time, our second intervention, communicating feedback results regularly with parents should be able to boost support especially for this group.

In fact, out results show that feedback can be a very effective method to raise test scores, both for left-behind and non-left-behind children. Effect sizes for math improvement is of the order of 0.20 standard deviations for both grades 3 and 5. For the left-behind group we find an additional 0.30 standard deviations improvement for the younger children (grade 3) when we bring the parents in via our second intervention, though this method has little impact for children with normal parenting. Overall, our feedback approach is indeed effective, and non-intrusive – and much cheaper than other methods such as intensive tutoring, or reducing class sizes.

Student feedback, parent-teacher communication, and academic performance: Experimental evidence from rural China

1. Introduction

Our study reports on randomized controlled trials designed to improve teacher-student-parent communication, conducted in a rural county in China with many left-behind children. The communication is based on a specially designed feedback scorecard. Our scorecard follows the classic feedback formula (Hattie, 2012) of providing objective information on the student's weekly academic progress and classroom behavior. At the same time, it is intended to be non-judgmental (Kluger and De Nisi, 1996) so that weaker students are not discouraged. In our first intervention, the results of our feedback form are discussed bi-weekly between teacher and student, and at the end of the year, we test for improvement in academic performance. In the second intervention, the results are also regularly shared with the parents, and again we test for improvement. Let us discuss these interventions.

First, we put forward our regular teacher-student feedback as a simple pedagogical mechanism for improving teachings standards. Alternatives such as high dosage tutoring (see Fryer, 2014) are expensive, and our simple feedback approach is worth trying. Hattie's (2012) well-known research, which examines 900 meta-analyses of what works best in education, gives feedback as one of the most potent influences. Effective feedback requires teacher judgment about a student's performance gap: ''Feedback is information about the gap between the actual level and the reference level of a system parameter, which is used to alter the gap in some way'' (Ramaprasad, 1983, cited in William, 2011, p4). Such feedback can be minimal "verification" or more extensive "evaluative" (e.g., Butler et al., 2013), and it can be directed at the task, the process or the student's "self" (Brown, 2012). The consensus developing is that more evaluative feedback is superior, with the UK's Assessment Reform group advocating movement from summative to formative assessment.[1] However, feedback that is more extensive of course requires more teacher time and effort, and there is not enough of it. For example, Pianta et al.'s (2007) large US study based on the National Institute of Child Health and Human Development dataset reports a low rating on the quality of evaluative feedback: "teachers gave generic feedback on correctness rather than encouraging extension of student performance or discussing alternative solutions". There is therefore a need to find more effective feedback methods, as in our simple scorecard approach.

At the same time, it is accepted (William, 2011) that the feedback gap must be negative (performance below target) for some, which can actually reduce performance if these students become discouraged. Thus, Kluger and DeNisi's (1996) well-cited analysis of feedback interventions found that while the average intervention had a large positive standard deviation effect size of 0.4, over one-third of the interventions actually decreased performance (see also Kosko and Miyazaki, 2012). In fact, while it is likely that students with high intrinsic motivation (i.e., more free time spent on the task – see Deci et al., 1999), or with high self-efficacy (Narciss, 2004) will face positive feedback gaps, raise their aspirations and move ahead, the converse may also be true. Moreover, research suggests (e.g. Hoover-Dempsey et al., 2001; Hill and Tyson, 2009) that student motivation and confidence are linked to parental support at home, e.g., with homework. Thus, we can expect the teacher-student feedback in our first intervention to work better for those living at home, and also with more educated or better-off parents who are more likely to provide the necessary support.

This argument underlines the problem that negative feedback comments are more likely for the left-behind children whose baseline math and Chinese test scores tend to be below the rest. These students are then more likely to be discouraged and change or abandon the goal, which is a bad result of course. However, enlisting parental support by communicating feedback results to

[1] There is also some US evidence (Hattie and Timperley, 2007, p100) that the better teachers, who are more "accomplished" on board certification assessments, use more feedback in the classroom. Also, Van Loon and Roebers's (2017) study of Swiss 4th-6th graders finds specific feedback helpful in reducing student overconfidence, which helps performance.

them (our second intervention below) is likely “to make academic goals salient” for the learners (Hattie and Timperley, 2007, p104), resulting in higher achievement. We thus expect that our intervention based on regular feedback will be effective when presented to the average learner who receives positive or at least not too bad feedback scores. But our second intervention, which brings the parents in, will help particularly for the left-behind or poor children who tend to receive negative feedback.

Our second intervention communicates the student’s feedback results to his/her parents, and can be seen as boosting parental education investments in the cognitive achievement production function (see Todd and Wolpin, 2003, and more recently Fredrikkson et al., 2016). In principle, parents and teachers should regularly communicate with each other to provide a supportive learning environment for schoolchildren (see e.g., Cunha and Heckman 2008l; Hill and Tyson, 2009; Cheung and Pomerantz, 2012). However adequate parent-teacher communication will be more difficult to achieve in poor communities where parents are less educated and informed about parenting techniques and child development (Kerr and Stattin, 2000; Bursztyn and Coffman, 2012). Furthermore, in China many parents need to work far from home and have to leave their children behind (see Wang and Mesman 2015 for a meta-analysis)[2]. Consequently, our intervention to share feedback results with parents should help the migrant parents most, since they face obstacles in productively investing in their children, for example, in monitoring homework (see Zhang et al 2014). In other words, we may have here another example of the empirical regularity noted by Fredrikkson et al. (2016), that the effect of education interventions is normally greatest for pupils from poorer families.

In recent years, there have been several well-designed experimental studies examining the effects of different parent-teacher communication programs on children’s educational outcomes.

[2] The phenomenon of large-scale worker migration and consequent left-behind children happens not only in China but also in many other developing countries, with negative impacts on the children (see Acosta (2011) on El Salvador, McKenzie and Rapoport (2011) on Mexico, and Cortes (2015) on the Philippines).

Avvisati et al. (2013) find that inviting parents in a deprived school district near Paris to attend three meetings with the school principal increases school-related parental care and induces positive behaviors and attitudes among Grade 6 students. Bergman (2015) finds that frequently informing parents in a low-income area of Los Angeles about their children's assignment quality for six months reduces parents' upwardly-biased belief about children's effort in school, raises parental monitoring and involvement, and improves behaviors and test scores. Kraft and Dougherty (2013) find that giving parents daily phone calls and messages improves homework completion and class participation among students in a mandatory summer school in Boston. In a similar study, Kraft and Rogers (2014) also find that delivering one-sentence individualized messages weekly to parents reduces dropouts among high school students in a credit recovery program. Islam (2017) finds that regular face-to-face meetings between parents and teachers in remote communities in rural Bangladesh increases parental involvement and student test scores. Lieberman et al. (2014), however, find that a more general parental education "information" program in rural Kenya produces no discernible effects on various student and parent behaviors.

While most of these experimental studies provide evidence on the positive effects of parent-teacher communication programs, we do not know how they interact with simply improving feedback as an instructional tool (our first intervention). There are two important sets of unaddressed concerns. First, it may be that a stand-alone student-teacher feedback program is sufficient, and (more costly) efforts towards enhancing parent-teacher communication bring little further benefit. Second, effects of student assessment and parent-teacher communication programs should be heterogeneous, as we have noted. In particular, we would expect the extra parent-teacher communication intervention to be of more assistance to children who are left behind, and more generally to children of lower baseline academic performance, in which case the extra expense is more likely to be worthwhile.

To answer these questions, we developed a special assessment scorecard, and tested its application in two randomized controlled trials (RCTs). The scorecard is shown in Figure 1, and as can be seen is simple, yet focused, and produces easily measurable results at a reasonable cost (we detail costs later). Our RCTs are based on results drawn from over 4,000 primary students in Shaoyang county, a low-income rural district in China's Hunan province. We were able to use administrative data on students and their examination results collected from the local education bureau, guaranteeing accuracy.

Our trials lasted for two school terms and included two experiments as noted. In the first, we asked class teachers to conduct student feedback assessments every two weeks[3] using our assessment scorecard. Teachers were then required to have individual meetings with their students to discuss results and offer advice. We regard this approach as testing teacher-student feedback (TSF). Our second experiment then built on the first one. Here, we required teachers of a randomly selected subset of students additionally to transmit the assessment results to the parents. Since many parents in our study sample worked far away from home, we asked teachers to use a simple smartphone application to transmit the results (see also Joshi et al., 2017). We regard this approach as testing teacher-student-parent communication (TSPC).[4]

The rest of this paper is organized as follows. Section 2 provides an overview of our research venue and background. Section 3 describes our research methodology, which includes the sampling and allocation methods, data collection, experimental interventions, and statistical approach. Section 4 presents the descriptive findings and the estimation results on the program effects of our interventions. Section 5 discusses and concludes.

---

[3] Bangert-Downs et al (1991) found that "class tests" occurring more frequently than every 2 weeks produced no extra benefit. For their part, Kluger and DeNisi's (1996) analysis suggested that a higher feedback frequency is better, without giving an actual figure.

[4] Our study does not include an experimental arm that contains only the parent-teacher communication program. This is because, as already mentioned above, student feedback assessment is essential to any individualized parent-teacher communication. Therefore, there is no way of asking teachers only to inform the parents about student feedback results without also talking to the students.

2. Research Venue and Background

We conducted our trials in 10 schools in Shaoyang county in China's Hunan Province (see Figure 1 for locations). Shaoyang county has a population of over one million, most of whom have a rural household registration (or a rural *Hukou*). The annual GDP per capita of the county is only around 13,000 yuan, which is roughly one-fourth of China's average. For many years, the county has been officially classified as a poverty county and has been receiving state-level anti-poverty supports. Notwithstanding such assistance, the county still faces a shortage of educational resources (e.g., low teacher-to-student ratios) and the educational performance of schoolchildren lags.

Like many other poor parts of rural China, a large number of parents in the county go to different cities along China's coastal areas to seek for better job opportunities and higher incomes. However, with a rural household registration, many of these parents choose not to bring their children with them because their children cannot receive public education and other public services outside their hometown (Chen and Feng, 2013). As a result, in 2016 for example, nearly 40 percent of the 61,000 primary schoolchildren and 50 percent of the 30,000 junior middle schoolchildren in the county had either one or both of their parents living away from home.[5]

In general, we would expect left-behind children generally to lag behind in both their cognitive and non-cognitive development, though the research findings are surprisingly variable (see e.g. Wang and Mesnan 2015; Ren and Treisman 2016). It seems to be important (see Zhang 2014) for both parents to be absent in order for negative effects to result – and our empirical procedures take this strict definition (i.e., both parents absent). It must also be remembered that left-behind children do not generally lack money, since the absent parents remit plentiful funds. For example, in Zhang et al's (2014) study in rural Hunan, families with both parents at home had only half the income of those with both parents absent. The problem is rather lack of parental knowledge and emotional support, even leading to depression amongst the children (He et al 2012;

[5] In 2013, in total through rural China, over 60 million children under the age of 17 had one or both of their parents migrating to the urban areas for work (All-China Women's Federation, 2013). On average, one in four children in China's rural areas were left-behind children. Hunan province, where our study is conducted, is among those provinces with the highest share, over 50 percent.

Shi et al 2016), and this effect becomes cumulatively worse (Meng and Yamauchi 2017). Naturally such children will have problems concentrating and remembering, and it is in precisely these circumstances that students find feedback uncomfortable (see Fyfe et al, 2015), suggesting that extra parental support as in our second intervention should be a good backup.[6] Indeed He et al (2015, p310) call for parents of left-behind children to seek ways to develop attachment relationships with their children which is exactly what our TSPC intervention does.

3. Research Methodology

We conducted our RCT in Shaoyang in four main stages. In the first stage (December 2014 to January 2015), we randomly selected 10 sample schools from Shaoyang county and collected basic information from administrative records on over 4,000 students. With the support of the local education bureau, we also were able to obtain the results of centrally administered uniform math and language (Chinese) tests.[7] These records provided our baseline characteristics and academic test scores for all the sample students. In the second stage after the winter break (February 2015 to March 2015), we prepared the schools for the random allocation protocol explained in more detail below. We then randomly allocated the students to the different experimental arms.

In the third stage of the RCT (from April 2015 to June 2015 and, after the summer break, from September 2015 to January 2016), we worked with the schools and the teachers to implement the two experiments. Finally, in the evaluation stage (the end of January 2016), we took advantage of another round of centrally administered uniform tests to obtain endline academic test scores for the sample students.

[6] Although China's central government has issued a set of policy guidelines to urge local governments to better protect these vulnerable children (State Council, 2016), local governments typically lack resources.

[7] In order to avoid possible observer/Hawthorne effects (e.g., Levitt and List 2011), we designed these exams as regular exams taking place in the end of the academic semester as normal, so the students were less likely to feel they were under particular observation.

3.1 Sampling and Allocation Methods

The first step of our sampling method is the selection of primary schools. In our study county (as well as in most rural counties of China), there are comprehensive primary schools and village primary schools. Comprehensive primary schools typically have all six primary school grades and there are multiple classes in each of the six grades. By contrast, village primary schools usually have only some of the six grades and there are only one or two classes per grade. We chose to conduct our study only in comprehensive primary schools because these schools are larger and better equipped with information and communication technologies (e.g., Internet-connected computers and Wifi networks) which are necessary for our parent-teacher intervention.[8] Also, since we wanted to allocate students to our experimental arms using a class-level randomization procedure, we needed sample schools that had multiple classes per grade. Overall, there were 25 comprehensive primary schools in our study county. We randomly selected 10 of them to constitute our sample (see Figure 3).[9]

In order to examine whether interventions are more effective for younger pupils, we targeted two school grades. Since the youngest Grade 1 and Grade 2 students would be unable to easily answer our survey questionnaire on personal characteristics, Grade 3 was chosen as the youngest group. In addition, we selected Grade 5 to represent the older group because Grade 4 students were just one year older and Grade 6 students would be under heavy pressure from public examinations for middle school places. In the 10 sample schools, we included all students in the two selected grades in our sample.

Since we could not treat students differently within classes, we randomized at class level. Ideally, we wanted to assign the same number of classes within each grade of each school to the three experimental arms. However, given variations in school size (some with only two classes per grade and some with more than three classes), we could not conduct a balanced assignment. Due

[8] In recent years the Chinese government has been merging village schools into comprehensive schools to better invest in school facilities (Liu et al., 2010; Wong et al., 2014).

[9] Seven of our ten sample schools are located in seven different towns or townships of the study county. The remaining three schools are located in the biggest town, which is also the county seat.

to resource constraints and also perceived fairness among schools, we randomly assigned one class in each grade of each school to be in the feedback classes arm of our study (TSF). In these classes, all students received biweekly feedback assessments from their class teachers. In addition, we then randomly assigned one class in each grade of the chosen schools to receive the parent-teacher communication program (TSPC)[10]. In these classes, all students received the feedback assessments and in addition, approximately one-third were randomly selected to have their bi-weekly results sent to their parents.

3.2 Experimental Interventions

Before carrying out the interventions, we held detailed training sessions with all class teachers in our sample schools and explained the implementation protocols. To encourage compliance and compensate for their additional workload, we provided the teachers a small monthly stipend. The local education bureau also issued to teachers who performed satisfactorily an official certificate to acknowledge their contribution. Throughout our intervention period, we also regularly visited the schools for monitoring purposes, and discussed with the principals, class teachers and students any implementation issues. For the parent-teacher communication intervention, we also made phone calls to some of the parents to check if they regularly received from teachers the feedback assessment results of their children. Overall, we found a high level of compliance.

The teacher-student feedback assessment (TSF). We worked with experienced teachers recruited from the county to develop a simple, non-judgmental feedback scorecard. As noted, this feedback concept is central to our project. The scorecard, shown in Figure 1, contains three categories: (1) academic performance (40 points maximum); (2) daily misconduct (90 points maximum); and (3) daily good behavior (15 bonus points maximum). To avoid discouraging poor performers, our form gives less weight to academic performance and more points to school behavior, which is

[10] However, to increase statistical power, in some of the larger schools we also randomly selected another class for this experimental arm.

easier to correct. We also attribute fewer points to good behavior than bad given the large class size in our sample schools (over 60 students per class on average), which makes good behavior harder to assess. In sum, the form helps teachers assess the academic performance and school behavior of the students in a simple yet comprehensive way. To reduce teacher subjectivity in the student assessments, before the intervention we conducted detailed training sessions with all class teachers and provided an instruction manual for reference.

The student feedback intervention lasted for two school terms. We asked the teachers to use the feedback form to assess their students every two weeks. After completing each round of the feedback, the teachers had to meet with the students individually for about ten minutes, and go through the results item-by-item. In this process, the teachers would also give students personalized feedback on how they could make improvements. Teachers completed five rounds in the first term and eight in the second term, for a total of thirteen feedback assessments.

Teacher-student-parent communication (TSPC). Our second intervention required communication between parents and teachers on the student feedback results using a smartphone application via *Wechat*[11], the most popular instant messaging smartphone application in China, which is free to use. After each feedback session, we asked teachers to take a photograph of the student's results with their smartphones, and then use *Wechat* to send the photos on to the parents. To ensure that parents could see the assessment results clearly, we gave teachers instructions on how to take good photographs of the assessment results. For monitoring purposes, teachers were also required to upload their photographs onto our confidential depository platform.

We designed this communication intervention as additional to the feedback process. In particular, we chose at random 50 (25) percent of the left-behind (non-left-behind) students receiving feedback to receive the communication intervention. Moreover, to help parents better

[11] This method is suitable given that over 75 percent of students' parents possess Internet-connected smartphones. Most of the other parents have smartphone access through relatives or coworkers. Also, all teachers in the sample classes had their own smartphones with school internet connection.

interpret the feedback results, we asked teachers to provide parents an explanatory note at the beginning of the process. Parents generally welcomed the assessments and many of them took the chance to have further conversation with the teachers. In sum, our intervention boosted two-way communication between parents and teachers .

3.3 Data Collection

The outcome measures of this study are test scores obtained from uniform math and language tests administered by the local education bureau. To make our empirical findings comparable with other studies, we convert raw scores obtained from these tests into standardized scores, calculated as below:

Standardized score = (Raw score – Average raw score in the control group) ÷ Standard deviation of raw scores in the control group.

We use the baseline average and standard deviation of the raw scores in the control group to standardize both the baseline and the endline test scores of all students by grades and by subjects. This is because the baseline raw scores in the control group could not be affected by the interventions.[12]

We then developed two intervention variables to represent the intervention status of our sample students. First, we constructed a binary variable, student feedback, to represent whether the student received biweekly feedback assessment from the class teachers. This variable equaled one (1) if the student was in a feedback class. The variable equaled zero if the student was in a control class. Then, we constructed another binary variable, parent-teacher communication, to represent whether the student received the parent-teacher communication intervention. This variable equaled

[12] As the baseline tests were administered before we allocated the sample students into different experimental arms, we also standardized the raw scores using the baseline scores of all sample students (i.e., all students in the three experimental arms). The results conducted with these standardized scores are essentially the same and are not shown for the sake of brevity.

one if the student was an assessment and communication class and was selected to receive the intervention, and zero otherwise.[13]

We also account for several student and teacher characteristics as control variables in our regression analysis. The student characteristics include gender, age (in months) and also left-behind status (both parents away from home=1; if not=0). In particular, for our empirical work, we define left-behind children strictly as those whose parents are both away from home, since as noted, it seems only the absence of both that really harms child development (Zhang et al., 2014). As regards teacher characteristics, we control for age, rank (i.e., senior teacher, level 1 teacher, level 2 teacher or others) and subject (i.e., teaching math or teaching language).[14]

3.4 Statistical Approach

We use a value-added model controlling for school dummy variables to estimate the causal effect of our interventions on academic performance:

$$Endline\ score_{ijk} = a_0 + a_1 TSF_{ijk} + a_2 TSPC_{ijk} + a_3 Baseline\ Score_{ijk} + a_4 X_i + \mu_k + e_{ijk}. \quad (1)$$

In the model, *i* represents students, *j* represents classes and *k* represents schools. The outcome variable is *Endline score* which represents the endline standardized test score for math or language. *Baseline score* is the corresponding baseline standardized test score and can be thought of as a good control for pre-existing student ability and family background factors. *TSF* and *TSPC* are the two intervention variables representing the intervention status of the student. We include student characteristics, X_i, to account for factors that may explain progress since the baseline, and confound results. School dummies, μ_k, are included in the regression to control for time-invariant

[13] In supplementary analysis, we find no spillover effect within the assessment and communication classes between students who were selected for the extra parent-teacher communication program and those who were not. Therefore, to better demonstrate the direct effect of our two programs, we define our intervention variables according to the actual intervention status of the students rather than the class assignment outcomes.
[14] Math and language are the two most important academic subjects in China's primary education and all of the class teachers in our sample taught either one of these two subjects.

heterogeneities at the school level. Given the RCT design, we hope that the estimates for a_1 and a_2 will tell us the causal effect of our interventions on student's academic performance.

We also refine model (1) to produce more robust estimates of the intervention effects given problems with balancing the control and treatment groups (discussed below). First, we further control for several teacher characteristics, X_j, in the regression model. Second, as the variations in school size constrained our experimental arm allocation strategy and affected the sample balance, we need a more flexible way of controlling for the baseline differences within school. Therefore, in the refined model, we interact school dummies, μ_k, with the baseline test score, *Baseline score*, to allow the important baseline performance variable to have a different effect for each school. The refined model is expressed as follows:

$$Endline\ score_{ijk} = a_0 + a_1 TSF_{ijk} + a_2 TSPC_{ijk} + a_3 Baseline\ Score_{ijk} + a_4 Baseline\ Score_{ijk} \times \mu_k + a_5 X_i + a_6 X_j + \mu_k + e_{ijk} \quad (2)$$

We conduct our regression analysis for the two school grades and the two academic subjects separately and use robust estimates of standard errors by clustering students within class.

4. Empirical Results

4.1 Descriptive Statistics

Table 1 presents the baseline distribution of classes and students in our 10 sample schools. As discussed above, all sample schools are comprehensive schools and most had between two and five classes per grade. The exception is School G which is a large central school located in the county seat (nine classes in Grade 3 and eight classes in Grade 5).

In total, as the last row shows, we randomly assigned 20 classes to the TSF intervention, which covered 1,133 students. We also randomly assigned 23 classes to be both feedback and communication classes. In these classes, 885 students received the TSF intervention and about one-third of them (469 students) were randomly selected to receive the additional parent-teacher

TSPC intervention.[15] Finally, the remaining 27 classes were not treated and served as controls. Thus, we had a baseline sample of over 2,000 students in 35 Grade 3 classes and over 2,000 students in another 35 Grade 5 classes.

However, because we faced such uneven school sizes in the county, our experimental arm allocation strategy was not able to achieve balance in the baseline characteristics of the sample. This problem is shown in Table 2. Specifically, the large central school G tends to have better results, and also contributed more students to the control group, so the baseline average scores for math and language tests were higher in the control group. In addition, since this school is located in the county seat and parents were more likely to find jobs locally, there were fewer left-behind children in the control group. Sample balance can be much improved if we exclude this school from the sample (as shown in Appendix Table 1), and we can also perform our analyses excluding this school. However, the results are little different when we do this (see Appendix Tables 7 and 8), and so we prefer to keep school G in our sample and more flexibly control for school differences in our regression framework (i.e., using model (2) as discussed above).

A further possible threat to our analysis is selection out of the sample. For various reasons (e.g., sick leaves, dropping out, changing schools, and migration to the city with parents), there was attrition of students over the study period. As Figure 3 shows, we were not able to follow up 370, or about 17% of the treatment students, compared to 220 or 13% of the controls, a difference of 4%. However, comparing the baseline characteristics of students in the final sample with those who were lost in our follow-up, we found no systematic difference except that the baseline math scores of the leavers is slightly higher (Appendix Table 3). This impact is small compared to the effect sizes that we estimate below, leaving our qualitative conclusions unchanged.

We start with a simple difference-in-difference (DID) approach to examine the effect of the two interventions. All estimates are presented in standard deviation units, σ. Table 3 presents

[15] In the assessment and communication classes we selected about 50 percent of left-behind children and 25 percent of the non-left-behind children to receive the additional TSPC intervention. Therefore, as shown in Appendix Table 2, within these classes, the left-behind status of the children is statistically associated with TSPC. However, we found no other statistical association between those who were selected to receive the intervention and those who were not.

initial evidence that the two interventions produce positive effects on academic performance. In Grade 3, the estimated effect of the student feedback intervention is 0.141σ for math scores and 0.096σ for language. In Grade 5, the DID estimates of the feedback effects are larger for both subjects (0.260σ for math and 0.204σ for language).

Combining student feedback with additional parent-teacher communication (TSPC) does not bring a significantly larger effect. This result is most easily seen in the last row of each panel which tests for the TSPC – TSF difference, which varies from positive to negative, but is never significant. However, given the limitations in our experimental arm allocation strategies, we should be careful in interpreting these simple estimates. In the following, we adopt a regression framework to allow inclusion of extra teacher and student controls.

4.2 Multivariate results

The Full sample. The multivariate results for the full sample confirm the strength of our student feedback TSF intervention. Table 4 reports the estimates for Grade 3, and we concentrate on columns (3) and (6) with the full set of controls. We see that TSF intervention significantly raises math scores by 0.16σ. It also significantly raises language scores, though only by half as much, 0.09σ. The extra parent-teacher communication TSPC intervention is not significant, though the combined effects of the two interventions are quite large for math, 0.23σ.

Much the same pattern is shown for the Grade 5 group presented in Table 5. Again, we see that the TSF intervention has strong effects, this time extending to language, with effect sizes in the region of 0.20σ. The TSPC intervention still tends to be insignificant, though this needs to be examined in the context of the left-behind children (where parental support needs more mobilization), as we do later.

As a robustness check, we add to model (2) interaction variables constructed by multiplying the set of school dummies with different student characteristics (i.e.., student age, gender and left-behind status). By including these interactions, we allow for even more flexible ways to control for how these student characteristics in different schools may affect our results. Fortunately, as reported in Appendix Table 4 results are similar to those reported in Tables 4 and 5.

Left-Behind Children versus Non-Left-Behind. Results for this important comparison are shown in Table 6. As explained at the outset, we would expect the extra parent-teacher communication intervention TSPC to be of more assistance to children who are left behind, and more generally to children of lower baseline academic performance, since better-off parents can make up for schooling inadequacies. The feedback intervention TSF, by itself, could demotivate left-behind children who have less intrinsic motivation.

Considering Grade 3 first, beneficial effects are indeed clear for the left-behind children. The TSF effect is high for math, 0.18σ, so there is no evidence of demotivation here. Moreover, the extra impact of TSPC on math achievement is 0.33σ, which is striking. For non-left-behind children, only the TSF program brings benefits, and only for math, 0.16σ. The TSPC program has small and insignificant effects, which is in the direction we expect since parental support is already available for this group. Overall, our two interventions together raise the math score of left-behind children by 0.51σ, and their language score by 0.18σ. These effects are much larger than for the non-left-behind group who only benefit from the feedback intervention. Thus, for these young children, we have another case of the empirical regularity (Fredrikkson et al 2016) that educational interventions are most help to pupils from poorer families, taking poor broadly in this context to mean lack of parental emotional contact.

For the older children, in Grade 5, the picture is rather the reverse, in that the non-left-behind benefit more than the left-behind, and the TSF program has most effect. Thus, for non-left-behind children, TSF effects are 0.24σ for math, and similar for language, while being insignificant for the left-behind. Indeed, there are signs of a demotivating effect of TSF for the left-behind children in math, -0.02σ. The TSPC program might again help these schoolchildren, since it has a positive effect, 0.14σ to 0.16σ, but this is insignificant. In general, it seems that our feedback intervention must be strengthened if it is to reach disadvantaged children at older ages, but we are on the right lines.

Baseline Test Performance. Next, we divide the sample into three tercile groups, from low to high baseline test performance. A low baseline score will be associated with low student ability or weak motivation, and also low parental education and income, all of which could cause negative feedback and lower the TSF impact. These schoolchildren should, however, benefit from the TSPC intervention in a similar way to the left-behind group.

However, as shown in Table 7, we find no clear patterns by baseline test performance. In Grade 3, TSF is more of a benefit to the math scores of students in the middle and the top-thirds (0.16σ and 0.18σ). TSPC, for its part, generally has a weak effect. It cannot be said to benefit the weaker students and provide a way of supplementing their likely low parental support.

In Grade 5, again evidence is mixed. TSF greatly benefits the math scores of students in the bottom-third by 0.36σ, which is contrary to the demotivation hypothesis. Also, contrary to our hypothesis that the TSPC program would help less able students more, it is the parents of the more able in the upper terciles who seem to take most advantage of the communication opportunities[16]. Nevertheless, whatever the route, our combined program effect sizes are respectable for grade 5, varying between 0.13σ and 0.42σ, suggesting again that our interventions are on the right lines.

Other heterogeneous treatment effects. Heterogeneity of treatment effects by gender is also of interest. There is evidence of gender differences in elementary school results (Hill et al 2007 show boys lag in reading particularly), and in study habits (Deming et al 2014), with girls having more intrinsic motivation. Girls could therefore respond more positively to extra feedback than boys

[16] In fact, during our monitoring trips, teachers of the assessment and communication classes generally told us that parents of the middle-third and top-third students typically responded more frequently and actively to the feedback assessment results sent by the class teachers. Parents of the bottom-third students usually appeared less concerned about their children's education and took more time to reply to the teacher's messages.

(see also Fryer 2014). At the same time, there is the specifically Chinese factor of son preference (e.g., Qian, 2008) which might lead to more parental care being directed to boys.

Our results are shown in Table 8, and we see indeed in Grade 3 that TSF effect sizes are larger for girls (0.23σ. or math and 0.10σ for language). For Grade 5, however, the pattern is almost reversed. Girls continue to benefit, particularly in language (0.20σ), and boys catch up with effect sizes 0.16 to 0.29σ. In general, we do not find marked differences by gender, but it might be that our interventions enable the boys to catch up more quickly as they go up the grades.

Finally, it is worth examining whether characteristics of the class teachers interact with the two interventions to produce heterogeneous program effects. Research (e.g. Clotfelter et al 2006, Kosko and Miyazaki 2012) suggests that more experienced teachers achieve better results, and may also use more feedback (Hattie and Timperley 2007). Specifically, we can analyze the intervention effects by the class teacher's age, rank, and teaching specialism.

We do not find any strong evidence that particular teacher expertise matters. As an example, we show results by teacher specialism in Table 9 (see also Appendix Tables 5 and 6). Teachers specializing in math may achieve higher TSF effects in math, and similarly language specialists in language. There is a hint of this process in Grade 3, where the language specialists have an effect size of 0.21σ in language, but only 0.15σ in math. However, in Grade 5 the difference goes the wrong way, with the language specialists achieving less in language (0.14σ) than in math (0.19σ). These results suggest that our feedback intervention is simple enough not to require special teacher expertise, which helps if the idea of extra feedback is to be taken up for practical education policy.

5. Conclusions

Our study provides experimental evidence of positive effects of a non-intrusive pedagogical intervention based on feedback. We use accurate administrative data collected from the local education bureau and academic test scores obtained from centrally administered uniform tests. We began with two questions: first, whether the feedback intervention (TSF) could be a useful method of bringing higher achievement, without discouragement effects described in the literature (e.g., William 2011). Second, whether the extra effort of communicating the feedback results to parents (TSPC) would boost parental support (e.g. as in Islam 2016), benefitting the performance particularly of left-behind children.

In answer to our first question, the simple TSF intervention significantly raises school achievement, with effects ranging from 0.09 (language) to 0.16σ (math) in Grade 3, and about 0.20σ for both math and language in Grade 5. However, there may be evidence of demotivation for the left-behind group in Grade 5, which shows signs of a negative effect of TSF for math – though bringing in the parents helps here (see below). To put these effects in perspective, Fryer's (2014, Table III) high dosage tutoring intervention in elementary schools, which costs nearly $2000 yearly per student produces similar effect sizes of around 0.20σ per year, but only for math (though effects are higher in secondary schools). William (2011, 7) reports effect sizes ranging from 0.14σ for weak feedback up to 0.56σ for strong formative assessment. Thus our simple feedback system, with effect sizes towards the middle of this range, compares well.

The answer to our second question, bringing the parents in via the TSPC intervention, differs for students with parents at home compared with those left-behind. Our hypothesis is that the TSPC intervention will have more effect for the left-behind group who lack immediate parental support, and broadly, this is what we find. In particular, the extra TSPC effect size (on top of \0.20σ for TSF) for Grade 3 left-behind children in math is large, 0.33σ. This large effect contrasts with generally weak effects for other groups. Our findings here seem to contrast with optimistic studies of improved communication among teachers and parents, particularly Islam's (2016) experiment in rural Bangladesh involving face-to-face meetings between teachers and parents, which gives large effect sizes of 0.22σ (math) and 0.36σ (English). However, the disadvantaged families of rural Bangladesh might almost be comparable with the disadvantaged families of the left-behind children in our samples, suggesting that there need be no contradiction.

In other words, low income or otherwise deprived families generally benefit from encouragements to parental support. This said, our TSPC program certainly does not hurt any group, and the question of whether it should be undertaken for better-off families as well therefore rests on its costs, to which we now turn.

We may draw the following policy implications. First, giving rural students regular simple academic and behavioral feedback assessments (whether or not their parents are involved) based on our type of scorecard appears to be very effective. Given the relatively low cost of our feedback intervention (only 0.33 USD per student per month mainly for the extra teacher compensation), we believe our results show that we have found a feasible and cost-effective way to improve educational outcomes for the rural areas in China as well as similar areas in developing countries in general.

Furthermore, our results indicate that keeping the parents of the left-behind children informed of their children's performance in school underpins better performance when the children are young. In fact, our finding in Grade 3, that the student feedback TSF intervention brings a smaller effect than the parent-teacher communication TSPC, implies that the input of the parental side is important for young children who are left-behind. The non-left-behind children provide a contrasting story. Here, the student feedback intervention alone can raise education outcomes but the parent-teacher intervention brings little added effect. In other words, where parenting is adequate, direct contact with the students themselves via organized feedback has the most effect (as in Wong et al's (2014) study of rural children's nutrition where direct child contact was more effective than communicating with parents). Therefore, the policy implication is that the extra effort of parent-teacher communication is needed mainly when the parents are absent and when the children are young, but otherwise (while desirable) is not so essential. This said, our simple feedback scorecard approach appears pedagogically effective, and it is low cost.

References

Acosta, P. (2011). School attendance, child labour, and remittances from international migration in El Salvador. *Journal of Development Studies*, 47, 913–936.

All-China Women's Federation. (2013). *Research report on left-behind children and migrant children in rural China [in Chinese].* All-China Women's Federation. Chongqing: Southwest China Normal University Press.

Avvisati, F., Gurgand, M., Guyon, N. and Maurin, E. (2013). Getting parents involved: A field experiment in deprived Schools. *Review of Economic Studies*, 81(1), 57–83.

Bangert-Drowns, R.L., Kulik, J.A. and Kulik, C-L.C. (1991). Effects of frequent classroom testing. *Journal of Educational Research*, 85(2), 89–99.

Brown, G., Harris, L.R. and Harnett, J. (2012). Teacher beliefs about feedback within an assessment for learning environment: Endorsement of improved learning over student well-being. *Teaching and Teacher Education*, 28(7), 968–978.

Butler A., Godbole, N. and Marsh, E.J. (2013). Explanation feedback is better than correct answer feedback for promoting transfer of learning. *Journal of Educational Psychology*, 105(2), 290–298.

Bergman, P. (2015). Parent-child information frictions and human capital investment: Evidence from a field experiment. *CESifo Working Paper Series No.5391*.

Bursztyn, L. and Coffman, L.C. (2012). The schooling decision: Family preferences, intergenerational conflict, and moral hazard in the Brazilian favelas. *Journal of Political Economy*, 120(3), 359–397.

Chang, H., Dong, X.Y. and Macphail, F. (2011). Labor migration and time use patterns of the left-behind children and elderly in rural China. *World Development*, 39 (12), 2199–2210.

Chen, Y. and Feng, S. (2013). Access to public schools and the education of migrant children in China. *China Economic Review*, 26(75), 75–88.

Cheung, C.S. and Pomerantz, E.M. (2012). Why does parents' involvement enhance children's achievement? The role of parent-oriented motivation. *Journal of Educational Psychology*, 104(3), 820–832.

Clotfelter, C.T., Ladd, H.F. and Vigdor, J.L. (2006). Teacher-student matching and the assessment of teacher effectiveness. *Journal of Human Resources*, 41(4), 778–82.

Cortes, P. (2015). The feminization of international migration and its effects on the children left behind: Evidence from the Philippines. *World Development*, 65, 62–78.

Cunha, F. and Heckman, J.J. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources*, 43(4), 739–782.

Deci, E., Koestner, R. and Ryan, R. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125: 621–668.

Deming, D., Hastings, J., Kane, T. and Staiger, D. (2014). School choice, school quality, and postsecondary attainment. *American Economic Review*, 104(3), 991–1013.

Fredrikkson, P., Öckert, B. and Oosterbeek, H. (2016). Parental responses to public investments in children. *Journal of Human Resources*, 51, 832–68.

Fryer, R. (2014). Injecting charter school best practices into traditional public schools: Evidence from field experiments. *Quarterly Journal of Economics*, 129, 1355–1407.

Fyfe, E.R., DeCaro, M.S. and Rittle-Johnson, B. (2015). When feedback is cognitively-demanding: The importance of working memory capacity. *Instructional Science*, 43(1), 73–91.

Hattie, J. (2012). Know Thy Impact. *Educational Leadership*, 70(1), 18–23.

Hattie, J. and Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112.

He B., Fan, J.. Liu N., Li, H. Wang, Y. Williams J., and Wong K (2012). Depression risk of ‘left-behind children’ in rural China, *Psychiatry Research,* 200, 306-312.

Hill, C.J., Bloom, H.S., Black, A.R. and Lipsey, M.W. (2007). Empirical benchmarks for interpreting effect sizes in research. *MDRC Working Paper*. Retrieved from http://www.mdrc.org/sites/default/files/full_84.pdf.

Hill, N. and Tyson, D. (2009). Parental involvement in middle school: A meta-analytic assessment of the strategies that promote achievement. *Developmental Psychology*, 45(3), 740–63.

Hoover-Dempsey, K.V., Battiato, A.C., Walker, J.M.T., Reed, R.P., DeJong, J.M. and Jones, K.P. (2001). Parental involvement in homework. *Educational Psychologist*, 36(3), 195–209.

Islam, A. (2017). Parent-teacher meetings and student outcomes: Evidence from a developing country. *Monash University Working Paper.*

Joshi, A., Generalla, J., Thompson, B., and Haidet, P. (2017). Facilitating the feedback process on a clinical clerkship using a smartphone application. *Academic Psychiatry*, 41(5), 651–55.

Kerr, M. and Stattin, H. (2000). What parents know, how they know it, and several forms of adolescent adjustment: Further support for a reinterpretation of monitoring. *Developmental Psychology*, 36(3), 366–380.

Kluger, A.N. and DeNisi, A.S. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254–84.

Kraft, M.A. and Dougherty, S.M. (2013). The effect of teacher-family communication on student engagement: Evidence from a randomized field experiment. *Journal of Research on Educational Effectiveness*, 6(3), 199–222.

Kraft, M.A. and Rogers, T. (2015). The underutilized potential of teacher-to-parent communication: Evidence from a field experiment. *Economics of Education Review*, 47, 49–63.

Kosko, K. and Miyazaki, Y. (2012). The effect of student discussion frequency on fifth-grade students' mathematics achievement in U.S. schools. *Journal of Experimental Education*, 80, 175–95.

Levitt, S.D. and List, J.A. (2011). Was there really a Hawthorne effect at the Hawthorne plant? An analysis of the original illumination experiments. *American Economic Journal: Applied Economics*, 3(1), 224–238.

Lieberman, E., Posner, D. and Tsai, L. (2014). Does information lead to more active citizenship? Evidence from an education intervention in rural Kenya. *World Development*, 60, 69–83.

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

McKenzie, D. and Rapoport, H. (2011). Can migration reduce educational attainment? Evidence from Mexico. *Journal of Population Economics*, 24(4), 1331–1358.

Meng X. and Yamauchi, C. (2017). Children of migrants: the cumulative impact of parental migration on children's education and health outcomes in China, *Demography,* 54, 1677-1714.

Narciss, S. (2004). The impact of informative tutoring feedback and self-efficacy on motivation and achievement in concept learning. *Experimental Psychology*, 51(3), 214–228.

Pianta, R.C., Belsky, J., Houts, R. and Morrison, R. (2007). Opportunities to learn in America's elementary classrooms. *Science*, 315(5820), 1795–1796.

Ren, Q.and Treisman D. (2016). The consequences of parental labor migration in China for children's emotional wellbeing, *Demography*, 58, 46-67.

Shi Y., Bai, Y., Shen, Y., Kenny, K., and Rozelle, S. (2016). Effects of parental migration on mental health of left-behind children: Evidence from Northwestern China, 24(3), 105-122.

State Council. (2016). *State Council Guidelines on the Protection of Left Behind Children* [in Chinese]. http://www.gov.cn/zhengce/content/2016-02/14/content_5041066.htm.

Todd, P. and Wolpin, K. (2003). On the specification and estimation of the production function for cognitive achievement. *Economic Journal*, 113, 3–33.

Van Loon, M.H. and Roebers, C.M. (2017). Effects of feedback on self-evaluations and self-regulation in elementary school. *Applied Cognitive Psychology*, 31(5), 508–519.
Wang, L. and J. Mesman (2015). Child development in the face of rural-to-urban migration in China: A meta-analytic review, *Psychological Scicence*¸ 10(6), 813-831.
William, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37, 3–14.
Wong, H.L., Shi, Y., Luo, R., Zhang, L. and Rozelle, S. (2014). Improving the health and education of elementary schoolchildren in rural China: Iron supplementation versus nutritional training for parents. *Journal of Development Studies*, 50, 502–519.
Zhang, H., Behrman, J.R., Fan, C.S., Wei, X. and Zhang, J. (2014). Does parental absence reduce cognitive achievements? Evidence from rural China. *Journal of Development Economics*, 111, 181–195.

Table 1: Distribution of classes and students by sample schools in Shaoyang County

Sample schools	Number of classes				Number of students				
	All classes	TSF classes	TSPC classes	Control classes	All students	Students in TSF alone	Students in TSPC classes		Control classes
							TSF only	TSPC as well	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Grade 3									
A	3	1	1	1	126	43	25	14	44
B	2	1	1	0	59	27	21	11	0
C	3	1	1	1	202	67	37	31	67
D	2	1	1	0	65	33	21	11	0
E	4	1	2	1	364	86	136	54	88
F	2	1	1	0	98	50	26	22	0
G	9	1	2	6	741	89	112	47	493
H	3	1	1	1	209	75	33	29	72
I	4	1	1	2	233	57	32	23	121
J	3	1	1	1	162	55	41	15	51
Subtotal	35	10	12	13	2259	582	485	257	936
Panel B: Grade 5									
A	3	1	1	1	160	52	35	19	54
B	2	1	1	0	66	31	24	11	0
C	3	1	1	1	151	62	36	25	28
D	2	1	1	0	36	18	14	4	0
E	5	1	1	3	425	79	60	23	263
F	2	1	1	0	99	50	33	16	0
G	8	1	1	6	611	71	50	18	472
H	3	1	1	1	204	72	35	26	71
I	4	1	2	1	257	62	76	55	64
J	3	1	1	1	161	54	37	15	55
Subtotal	35	10	11	14	2170	551	400	212	1007
Total	70	20	23	27	4429	1133	885	469	1943

Table 2: Baseline characteristics of sample students across experimental arms

	Students receiving TSF [a]	Students receiving TSPC	Students in control classes	(1)-(3)	(2)-(3)
	(1)	(2)	(3)	(4)	(5)
			Panel A: Grade 3		
Standardized math scores	-0.265	-0.297	0.000	-0.265***	-0.297***
				(0.058)	(0.073)
Standardized language scores	-0.382	-0.284	-0.000	-0.382***	-0.284***
				(0.061)	(0.073)
Age (in months)	108.273	108.233	107.244	1.029*	0.990
				(0.416)	(0.544)
Female (1=Yes; 0=No)	0.457	0.463	0.455	0.001	-0.007
				(0.028)	(0.035)
Left-behind (1=Yes; 0=No)	0.180	0.381	0.218	-0.038*	0.163***
				(0.023)	(0.030)
			Panel B: Grade 5		
Standardized math scores	-0.591	-0.615	-0.000	-0.591***	-0.615***
				(0.066)	(0.081)
Standardized language scores	-0.410	-0.398	-0.000	-0.410***	-0.398***
				(0.064)	(0.080)
Age (in months)	132.275	132.995	131.352	0.923	1.644**
				(0.473)	(0.621)
Female (1=Yes; 0=No)	0.455	0.429	0.436	0.019	-0.007
				(0.029)	(0.037)
Left-behind (1=Yes; 0=No)	0.175	0.354	0.149	0.026	0.205***
				(0.021)	(0.029)

Note: Robust standard errors adjusted for clustering at the school level are reported in parentheses. ***, ** and * indicate statistical significance at 1, 5 and 10 percent levels.
[a] This column includes all students receiving the TSF treatment.

Table 3: Simple DID estimates of student assessment and parent-teacher communication on standardized math and language scores

	Standardized math scores			Standardized language scores		
	Baseline test	Endline test	Difference (2) – (1)	Baseline test	Endline test	Difference (5) – (4)
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Grade 3						
Teacher-Student feedback (TSF)[a]	-0.266	-0.130	0.136***	-0.369	-0.380	-0.011
			(0.053)			(0.053)
Teacher-Student-Parent communication (TSPC)	-0.281	-0.046	0.234**	-0.316	-0.333	-0.017
			(0.103)			(0.108)
Control group	0.000	-0.005	-0.005	0.000	-0.107	-0.107**
			(0.049)			(0.047)
Difference: TSF – Control			0.141*			0.096
			(0.079)			(0.067)
Difference: TSPC – Control			0.239**			0.090
			(0.093)			(0.061)
Difference: TSPC – TSF			0.098			-0.006
			(0.061)			(0.053)
Panel B: Grade 5						
Teacher-Student feedback (TSF)[a]	-0.770	-0.320	0.450***	-0.606	-1.174	-0.568***
			(0.073)			(0.066)
Teacher-Student-Parent communication (TSPC)	-0.763	-0.386	0.377**	-0.414	-0.920	-0.506***
			(0.154)			(0.130)
Control groups	0.000	0.189	0.189***	0.000	-0.772	-0.772***
			(0.048)			(0.047)
Difference: TSF – Control			0.260*			0.204
			(0.136)			(0.148)
Difference: TSPC – Control			0.188			0.266**
			(0.106)			(0.102)
Difference: TSPC – TSF			-0.072			0.062
			(0.099)			(0.138)

Note: Robust standard errors adjusted for clustering at the school level are reported in parentheses. ***, ** and * indicate statistical significance at 1, 5 and 10 percent levels.

[a] This column includes all students receiving the TSF treatment.

Table 4: Effects of student assessment and parent-teacher communication on standardized test scores of grade 3 students

	Dependent variable: Endline standardized math scores (standard deviation units)			Dependent variable: Endline standardized language scores (standard deviation units)		
	(1)	(2)	(3)	(4)	(5)	(6)
TSF (1=Students assessed biweekly; 0=Otherwise)	0.16**	0.17***	0.16**	0.07	0.10*	0.09*
	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)	(0.05)
TSPC (1=Parents received assessment results; 0=Otherwise)	0.09*	0.08	0.07	-0.00	-0.00	0.00
	(0.05)	(0.05)	(0.05)	(0.04)	(0.05)	(0.05)
Baseline standardized math scores	0.69***	0.69***	0.78***			
	(0.03)	(0.03)	(0.03)			
Baseline standardized language scores				0.66***	0.67***	0.72***
				(0.02)	(0.02)	(0.00)
Age (in months)	-0.01***	-0.01***	-0.01***	-0.00*	-0.00*	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Female (1=Yes; 0=No)	-0.02	-0.02	-0.02	0.21***	0.21***	0.20***
	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)
Left-behind (1=Both parents away; 0=Otherwise)	-0.05	-0.06	-0.05	0.02	0.02	0.02
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Teacher characteristics	N	Y	Y	N	Y	Y
School dummies	Y	Y	Y	Y	Y	Y
School dummies * Baseline test scores	N	N	Y	N	N	Y
N	1948	1948	1948	1936	1936	1936
R^2	0.58	0.58	0.60	0.70	0.71	0.72
Combined program effect	0.25***	0.25***	0.23***	0.06	0.09	0.09
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)

Note: Teacher characteristics include teacher's age (in month), teacher's rank (senior, level 1, level 2 or others), and whether the subject was taught by class teacher (1=Yes; 0=No). Robust standard errors adjusted for clustering at the class level are reported in parentheses. ***, **, and * indicate significance at the 1, 5 and 10 percent levels.

Table 5: Effects of student assessment and parent-teacher communication on standardized test scores of grade 5 students

	Dependent variable: Endline standardized math scores (standard deviation units)			Dependent variable: Endline standardized language scores (standard deviation units)		
	(1)	(2)	(3)	(4)	(5)	(6)
TSF (1=Students assessed biweekly; 0=Otherwise)	0.08	0.18***	0.20***	0.11	0.17*	0.18*
	(0.08)	(0.06)	(0.06)	(0.07)	(0.09)	(0.09)
TSPC (1=Parents received assessment results; 0=Otherwise)	0.04	-0.01	-0.01	0.18*	0.11	0.14
	(0.10)	(0.10)	(0.09)	(0.10)	(0.09)	(0.09)
Baseline standardized math scores	0.63***	0.64***	0.85***			
	(0.04)	(0.04)	(0.04)			
Baseline standardized language scores				0.56***	0.57***	0.64***
				(0.04)	(0.04)	(0.11)
Age (in months)	-0.01***	-0.01***	-0.01***	-0.00*	-0.00*	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Female (1=Yes; 0=No)	-0.02	-0.02	-0.01	0.32***	0.32***	0.31***
	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	(0.04)
Left-behind (1=Both parents away; 0=Otherwise)	-0.13*	-0.12*	-0.10	-0.04	-0.04	-0.03
	(0.07)	(0.07)	(0.06)	(0.03)	(0.03)	(0.03)
Teacher characteristics	N	Y	Y	N	Y	Y
School dummies	Y	Y	Y	Y	Y	Y
School dummies * Baseline test scores	N	N	Y	N	N	Y
N	1891	1891	1891	1891	1891	1891
R^2	0.62	0.63	0.65	0.62	0.63	0.65
Combined program effect	0.12	0.16	0.19**	0.29**	0.28**	0.31***
	(0.11)	(0.10)	(0.09)	(0.11)	(0.11)	(0.11)

Note: Teacher characteristics include teacher's age (in month), teacher's rank (senior, level 1, level 2 or others), and whether the subject was taught by class teacher (1=Yes; 0=No). Robust standard errors adjusted for clustering at the class level are reported in parentheses. ***, **, and * indicate significance at the 1, 5 and 10 percent levels.

Table 6: Effects of student assessment and parent-teacher communication by student's left-behind status

	Dependent variable: Endline standardized scores (standard deviation units)							
	Grade 3				Grade 5			
	Left-behind		Non left-behind		Left-behind		Non left-behind	
	Math	Lang.	Math	Lang.	Math	Lang.	Math	Lang.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
TSF (1=Students assessed biweekly; 0=Otherwise)	0.18* (0.09)	0.13 (0.10)	0.16** (0.07)	0.07 (0.05)	-0.02 (0.12)	0.14 (0.12)	0.24*** (0.06)	0.18** (0.09)
TSPC (1=Parents received assessment results; 0=Otherwise)	0.33*** (0.11)	0.05 (0.09)	-0.03 (0.07)	0.03 (0.05)	0.14 (0.15)	0.16 (0.16)	-0.07 (0.09)	0.15* (0.09)
Student and teacher characteristics	Y	Y	Y	Y	Y	Y	Y	Y
School dummies	Y	Y	Y	Y	Y	Y	Y	Y
School dummies * Baseline test scores	Y	Y	Y	Y	Y	Y	Y	Y
N	453	450	1495	1486	363	363	1528	1528
R^2	0.59	0.76	0.60	0.71	0.70	0.70	0.61	0.64
Combined program effect	0.51*** (0.12)	0.18* (0.10)	0.13 (0.08)	0.09 (0.07)	0.12 (0.17)	0.30* (0.16)	0.17* (0.09)	0.33*** (0.11)

Note: Student characteristics include age (in months), gender (female), and baseline standardized test scores of subjects. Teacher characteristics include teacher's age (in month), teacher's rank (senior, level 1, level 2 or others), and whether the subject was taught by class teacher (1=Yes; 0=No). Robust standard errors adjusted for clustering at the class level are reported in parentheses.
***, **, and * indicate significance at the 1, 5 and 10 percent levels.

Table 7: Effects of student assessment and parent-teacher communication by baseline academic performance

	Dependent variable: Endline standardized scores (standard deviation units)					
	Bottom third		Middle third		Top third	
	Math	Language	Math	Language	Math	Language
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Grade 3			
TSF (1=Students assessed biweekly; 0=Otherwise)	0.14	0.11*	0.16**	0.14**	0.18**	0.03
	(0.10)	(0.07)	(0.07)	(0.06)	(0.06)	(0.05)
TSPC (1=Parents received assessment results; 0=Otherwise)	0.12	-0.08	0.04	0.06	0.08*	0.07
	(0.15)	(0.07)	(0.07)	(0.09)	(0.04)	(0.06)
Student and teacher characteristics	Y	Y	Y	Y	Y	Y
School dummies	Y	Y	Y	Y	Y	Y
School dummies * Baseline test scores	Y	Y	Y	Y	Y	Y
N	646	628	632	641	670	667
R^2	0.49	0.63	0.19	0.21	0.17	0.20
Combined program effect	0.26*	0.03	0.21**	0.20*	0.26***	0.11
	(0.16)	(0.09)	(0.09)	(0.10)	(0.06)	(0.09)
			Panel B: Grade 5			
TSF (1=Students assessed biweekly; 0=Otherwise)	0.36***	0.08	0.08	0.20	0.06*	0.31**
	(0.11)	(0.09)	(0.12)	(0.13)	(0.03)	(0.12)
TSPC (1=Parents received assessment results; 0=Otherwise)	-0.08	0.06	0.07	0.21	0.07*	0.10
	(0.15)	(0.13)	(0.17)	(0.15)	(0.04)	(0.09)
Student and teacher characteristics	Y	Y	Y	Y	Y	Y
School dummies	Y	Y	Y	Y	Y	Y
School dummies * Baseline test scores	Y	Y	Y	Y	Y	Y
N	627	618	563	641	701	632
R^2	0.52	0.58	0.28	0.22	0.20	0.31
Combined program effect	0.28*	0.13	0.14	0.42**	0.13***	0.41**
	(0.16)	(0.14)	(0.17)	(0.18)	(0.04)	(0.12)

Note: Student and teacher characteristics are listed in Table 6. Robust standard errors adjusted for clustering at the class level are reported in parentheses. ***, **, and * indicate significance at the 1, 5 and 10 percent levels.

Table 8: Effects of student assessment and parent-teacher communication by student's gender

	Dependent variable: Endline standardized scores (standard deviation units)							
	Grade 3				Grade 5			
	Girls		Boys		Girls		Boys	
	Math	Lang.	Math	Lang.	Math	Lang.	Math	Lang.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
TSF (1=Students assessed biweekly; 0=Otherwise)	0.23***	0.10**	0.09	0.06	0.06	0.20**	0.29***	0.16
	(0.07)	(0.04)	(0.09)	(0.05)	(0.07)	(0.09)	(0.07)	(0.11)
TSPC (1=Parents received assessment results; 0=Otherwise)	0.02	-0.07	0.09	0.07	0.06	0.14*	-0.06	0.13
	(0.09)	(0.06)	(0.07)	(0.06)	(0.12)	(0.08)	(0.13)	(0.12)
Student and teacher characteristics	Y	Y	Y	Y	Y	Y	Y	Y
School dummies	Y	Y	Y	Y	Y	Y	Y	Y
School dummies * Baseline test scores	Y	Y	Y	Y	Y	Y	Y	Y
N	871	867	1077	1069	831	831	1060	1060
R^2	0.65	0.67	0.57	0.72	0.66	0.62	0.66	0.63
Combined program effect	0.24**	0.02	0.17*	0.13*	0.12	0.34***	0.23*	0.29**
	(0.09)	(0.07)	(0.09)	(0.07)	(0.12)	(0.11)	(0.13)	(0.13)

Note: Student and teacher characteristics are listed in Table 6. Robust standard errors adjusted for clustering at the class level are reported in parentheses. ***, **, and * indicate significance at the 1, 5 and 10 percent levels.

Table 9: Effects of student assessment and parent-teacher communication by the teaching subject of class teachers

	Dependent variable: Endline standardized scores(standard deviation units)							
	Grade 3				Grade 5			
	Math teacher		Language teacher		Math teacher		Language teacher	
	Math	Lang.	Math	Lang.	Math	Lang.	Math	Lang.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
TSF (1=Students assessed biweekly; 0=Otherwise)	0.18* (0.09)	0.18 (0.16)	0.15* (0.07)	0.21*** (0.05)	0.45*** (0.11)	0.44 (0.29)	0.19** (0.07)	0.14 (0.11)
TSPC (1=Parents received assessment results; 0=Otherwise)	0.07 (0.08)	0.07 (0.05)	0.13 (0.09)	-0.07 (0.07)	0.08 (0.12)	0.21 (0.13)	-0.14 (0.13)	0.10 (0.10)
Student and teacher characteristics	Y	Y	Y	Y	Y	Y	Y	Y
School dummies	Y	Y	Y	Y	Y	Y	Y	Y
School dummies * Baseline test scores	Y	Y	Y	Y	Y	Y	Y	Y
N	775	770	1173	1166	571	571	1320	1320
R^2	0.52	0.70	0.63	0.74	0.69	0.72	0.64	0.60
Combined program effect	0.26** (0.10)	0.25 (0.15)	0.27*** (0.09)	0.14* (0.07)	0.53*** (0.09)	0.64** (0.29)	0.05 (0.12)	0.24 (0.14)

Note: Student and teacher characteristics are listed in Table 6. Robust standard errors adjusted for clustering at the class level are reported in parentheses. ***, **, and * indicate significance at the 1, 5 and 10 percent levels.

Appendix Table 1: Baseline characteristics of sample students across experimental arms without School G

	Students receiving TSF [a]	Students receiving TSPC	Students in control classes	(1)-(3)	(2)-(3)
	(1)	(2)	(3)		
			Panel A: Grade 3		
Standardized math scores	-0.038	-0.034	-0.000	-0.038	-0.034
				(0.071)	(0.086)
Standardized language scores	-0.272	-0.141	-0.000	-0.272***	-0.141
				(0.076)	(0.086)
Age (in months)	108.677	108.476	107.828	0.849	0.648
				(0.600)	(0.761)
Female (1=Yes; 0=No)	0.452	0.462	0.474	-0.022	-0.012
				(0.035)	(0.042)
Left-behind (1=Yes; 0=No)	0.215	0.433	0.375	-0.160***	0.059
				(0.032)	(0.041)
			Panel B: Grade 5		
Standardized math scores	-0.288	-0.283	-0.000	-0.288***	-0.283**
				(0.075)	(0.088)
Standardized language scores	-0.058	-0.015	-0.000	-0.058	-0.015
				(0.074)	(0.088)
Age (in months)	132.574	133.062	132.350	0.225	0.712
				(0.575)	(0.730)
Female (1=Yes; 0=No)	0.454	0.438	0.437	0.017	0.001
				(0.034)	(0.042)
Left-behind (1=Yes; 0=No)	0.197	0.381	0.228	-0.031	0.153***
				(0.028)	(0.037)

Notes: Robust standard errors adjusted for clustering at the school level are reported in parentheses. ***, ** and * indicate statistical significance at 1, 5 and 10 percent levels.
[a] This column includes all students in the feedback assessment classes and students in the feedback assessment and communication classes who received student feedback only.

Appendix Table 2: Baseline characteristics of sample students in assessment and communication classes

	Students in assessment and communication classes		(1)-(2)
	Receiving TSF only	Receiving TSPC	
	(1)	(2)	(3)
		Panel A: Grade 3	
Standardized math scores	-0.247	-0.289	0.041(0.087)
Standardized language scores	-0.384	-0.256	-0.128(0.093)
Age (in months)	108.135	108.195	-0.060(0.520)
Female (1=Yes; 0=No)	0.468	0.463	0.004(0.039)
Left-behind (1=Yes; 0=No)	0.206	0.354	-0.148***(0.034)
N	496	246	742
		Panel B: Grade 5	
Standardized math scores	-0.604	-0.621	0.017(0.115)
Standardized language scores	-0.462	-0.410	-0.052(0.111)
Age (in months)	132.282	133.125	-0.843(0.750)
Female (1=Yes; 0=No)	0.453	0.423	0.030(0.042)
Left-behind (1=Yes; 0=No)	0.198	0.351	-0.153***(0.036)
N	404	208	612

Note: Robust standard errors adjusted for clustering at the school level are reported in parentheses. ***, ** and * indicate statistical significance at 1, 5 and 10 percent levels.

Appendix Table 3: Attrition analysis of sample students

	Dependent variable: Student missing in the endline math test (1=Yes; 0=No)			Dependent variable: Student missing in the endline language test (1=Yes; 0=No)		
	(1)	(2)	(3)	(4)	(5)	(6)
Student in TSF classes (1=Yes; 0=No)	0.01	0.01	0.02	0.01	0.01	0.02
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Student in TSPC classes (1=Yes; 0=No)	-0.00	-0.00	-0.02	-0.00	-0.00	-0.02
	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)
Grade 5 (1=Yes; 0=No)	-0.02	-0.00	-0.00	-0.01	-0.01	0.00
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Baseline standardized math scores	0.02***	0.02***	0.05***			
	(0.01)	(0.01)	(0.01)			
Baseline standardized language scores				0.01	0.01	-0.00
				(0.01)	(0.01)	(0.02)
Age (in months)		0.00	0.00		0.00	0.00
		(0.00)	(0.00)		(0.00)	(0.00)
Female (1=Yes; 0=No)		0.01	0.01		0.01	0.01
		(0.01)	(0.01)		(0.01)	(0.01)
Left-behind (1=Both parents away; 0=Otherwise)		-0.01	-0.01		-0.01	-0.01
		(0.01)	(0.01)		(0.01)	(0.01)
Teacher characteristics	N	N	Y	N	N	Y
School dummies	Y	Y	Y	Y	Y	Y
School dummies * Baseline test scores	N	N	Y	N	N	Y
N	4429	4429	4429	4429	4429	4429
R^2	0.15	0.16	0.18	0.15	0.15	0.17

Note: Student and teacher characteristics are listed in Table 6. Robust standard errors adjusted for clustering at the class level are reported in parentheses. ***, **, and * indicate significance at the 1, 5 and 10 percent levels.

Appendix Table 4: Robustness check of the effects of student assessment and parent-teacher communication

	Dependent variable: Endline standardized math scores (std. dev. units)	Dependent variable: Endline standardized lang. scores (std. dev. units)
	Panel A: Grade 3	
TSF (1=Students assessed biweekly; 0=Otherwise)	0.16**	0.08*
	(0.06)	(0.05)
TSPC (1=Parents received assessment results; 0=Otherwise)	0.07	0.02
	(0.05)	(0.05)
Student and teacher characteristics	Y	Y
School dummies	Y	Y
School dummies * Baseline test scores	Y	Y
School dummies * Age	Y	Y
School dummies * Female	Y	Y
School dummies * Left-Behind	Y	Y
N	1948	1936
R^2	0.60	0.72
Combined program effect	0.23***	0.09
	(0.06)	(0.06)
	Panel b: Grade 5	
TSF (1=Students assessed biweekly; 0=Otherwise)	0.20***	0.18*
	(0.05)	(0.09)
TSPC (1=Parents received assessment results; 0=Otherwise)	-0.00	0.13
	(0.10)	(0.09)
Student and teacher characteristics	Y	Y
School dummies	Y	Y
School dummies * Baseline test scores	Y	Y
School dummies * Age	Y	Y
School dummies * Female	Y	Y
School dummies * Left-Behind	Y	Y
N	1891	1891
R^2	0.66	0.66
Combined program effect	0.19**	0.32***
	(0.10)	(0.11)

Note: Student and teacher characteristics are listed in Table 6. Robust standard errors adjusted for clustering at the class level are reported in parentheses. ***, **, and * indicate significance at the 1, 5 and 10 percent levels.

Appendix Table 5: Effects of student assessment and parent-teacher communication by teacher's age

	Dependent variable: Endline standardized scores (standard deviation units)							
	Grade 3				Grade 5			
	Young teachers		Old teachers		Young teachers		Old teachers	
	Math	Lang.	Math	Lang.	Math	Lang.	Math	Lang.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
TSF (1=Students assessed biweekly; 0=Otherwise)	0.19***	0.23**	0.26**	0.21**	0.15**	0.30***	0.22**	0.10
	(0.06)	(0.09)	(0.10)	(0.09)	(0.05)	(0.08)	(0.09)	(0.13)
TSPC (1=Parents received assessment results; 0=Otherwise)	0.06	-0.05	0.09	0.14	-0.10	0.02	0.04	0.15
	(0.06)	(0.06)	(0.10)	(0.08)	(0.14)	(0.07)	(0.12)	(0.12)
Student and teacher characteristics	Y	Y	Y	Y	Y	Y	Y	Y
School dummies	Y	Y	Y	Y	Y	Y	Y	Y
School dummies * Baseline test scores	Y	Y	Y	Y	Y	Y	Y	Y
N	1020	1019	928	917	793	793	1098	1098
R^2	0.64	0.76	0.57	0.68	0.67	0.66	0.64	0.69
Combined program effect	0.25***	0.17*	0.35***	0.34**	0.05	0.27**	0.26**	0.25
	(0.06)	(0.08)	(0.11)	(0.14)	(0.12)	(0.11)	(0.12)	(0.17)

Note: Student and teacher characteristics are listed in Table 6. Robust standard errors adjusted for clustering at the class level are reported in parentheses. ***, **, and * indicate significance at the 1, 5 and 10 percent levels.

Appendix Table 6: Effects of student assessment and parent-teacher communication by teacher's rank

	Dependent variable: Endline standardized scores (standard deviation units)							
	Grade 3				Grade 5			
	Junior rank teachers		Senior rank teachers		Junior rank teachers		Senior rank teachers	
	Math	Lang.	Math	Lang.	Math	Lang.	Math	Lang.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
TSF (1=Students assessed biweekly; 0=Otherwise)	0.15 (0.11)	0.03 (0.05)	0.24*** (0.06)	0.28*** (0.09)	0.25* (0.14)	0.42 (0.25)	0.18** (0.08)	-0.01 (0.13)
TSPC (1=Parents received assessment results; 0=Otherwise)	0.03 (0.06)	0.02 (0.06)	0.12 (0.08)	-0.03 (0.07)	-0.07 (0.12)	-0.02 (0.06)	-0.00 (0.14)	0.23* (0.12)
Student and teacher characteristics	Y	Y	Y	Y	Y	Y	Y	Y
School dummies	Y	Y	Y	Y	Y	Y	Y	Y
School dummies * Baseline test scores	Y	Y	Y	Y	Y	Y	Y	Y
N	915	915	1033	1021	703	703	1188	1188
R^2	0.61	0.73	0.60	0.72	0.67	0.65	0.64	0.68
Combined program effect	0.18 (0.12)	0.05 (0.07)	0.37*** (0.08)	0.25** (0.11)	0.19 (0.11)	0.41 (0.26)	0.18 (0.14)	0.22 (0.19)

Note: Student and teacher characteristics are listed in Table 6. Robust standard errors adjusted for clustering at the class level are reported in parentheses. ***, **, and * indicate significance at the 1, 5 and 10 percent levels.

Appendix Table 7: Effects of student assessment and parent-teacher communication on standardized test scores of grade 3 students (without school G)

	Dependent variable: Endline standardized math scores (standard deviation units)			Dependent variable: Endline standardized language scores (standard deviation units)		
	(1)	(2)	(3)	(4)	(5)	(6)
TSF (1=Students feedback biweekly; 0=Otherwise)	0.14*	0.11	0.10	0.14**	0.13**	0.13**
	(0.07)	(0.07)	(0.07)	(0.06)	(0.06)	(0.06)
TSPC (1=Parents received feedback results; 0=Otherwise)	0.09	0.07	0.07	-0.02	-0.03	-0.02
	(0.06)	(0.07)	(0.07)	(0.05)	(0.06)	(0.06)
Baseline standardized math scores	0.74***	0.74***	0.78***			
	(0.03)	(0.03)	(0.03)			
Baseline standardized language scores				0.71***	0.71***	0.72***
				(0.02)	(0.01)	(0.01)
Age (in months)	-0.01***	-0.01***	-0.01***	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Female (1=Yes; 0=No)	-0.01	-0.01	-0.00	0.18***	0.18***	0.18***
	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)
Left-behind (1=Both parents away; 0=Otherwise)	-0.07	-0.07	-0.07	0.00	-0.00	-0.00
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Teacher characteristics	N	Y	Y	N	Y	Y
School dummies	Y	Y	Y	Y	Y	Y
School dummies * Baseline test scores	N	N	Y	N	N	Y
N	1243	1243	1243	1242	1242	1242
R^2	0.59	0.59	0.61	0.72	0.72	0.73
Combined program effect	0.23***	0.18*	0.17	0.12*	0.11	0.10
	(0.07)	(0.09)	(0.09)	(0.06)	(0.08)	(0.08)

Note: Student and teacher characteristics are listed in Table 6. Robust standard errors adjusted for clustering at the class level are reported in parentheses. ***, **, and * indicate significance at the 1, 5 and 10 percent levels.

Appendix Table 8: Effects of student assessment and parent-teacher communication on standardized test scores of grade 5 students (without school G)

	Dependent variable: Endline standardized math scores (standard deviation units)			Dependent variable: Endline standardized language scores (standard deviation units)		
	(1)	(2)	(3)	(4)	(5)	(6)
TSF (1=Students assessed biweekly; 0=Otherwise)	0.09	0.18**	0.20***	0.06	0.16*	0.19*
	(0.10)	(0.06)	(0.06)	(0.09)	(0.09)	(0.09)
TSPC (1=Parents received assessment results; 0=Otherwise)	0.04	-0.04	-0.04	0.21*	0.11	0.13
	(0.11)	(0.12)	(0.11)	(0.12)	(0.10)	(0.10)
Baseline standardized math scores	0.64***	0.65***	0.84***			
	(0.04)	(0.04)	(0.04)			
Baseline standardized language scores				0.57***	0.57***	0.65***
				(0.04)	(0.04)	(0.11)
Age (in months)	-0.01***	-0.01***	-0.01***	-0.01*	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Female (1=Yes; 0=No)	0.02	0.02	0.02	0.39***	0.39***	0.38***
	(0.05)	(0.05)	(0.05)	(0.07)	(0.07)	(0.06)
Left-behind (1=Both parents away; 0=Otherwise)	-0.16*	-0.15*	-0.13*	-0.04	-0.04	-0.04
	(0.08)	(0.08)	(0.07)	(0.03)	(0.03)	(0.03)
Teacher characteristics	N	Y	Y	N	Y	Y
School dummies	Y	Y	Y	Y	Y	Y
School dummies * Baseline test scores	N	N	Y	N	N	Y
N	1305	1305	1305	1304	1304	1304
R^2	0.60	0.60	0.63	0.60	0.62	0.64
Combined program effect	0.13	0.14	0.17	0.27**	0.28**	0.32***
	(0.13)	(0.11)	(0.11)	(0.12)	(0.12)	(0.12)

Note: Student and teacher characteristics are listed in Table 6. Robust standard errors adjusted for clustering at the class level are reported in parentheses. ***, **, and * indicate significance at the 1, 5 and 10 percent levels.

Assessment categories	Assessment items	Points	Remarks
Academic performance (40 points maximum)	Finishing homework on time		(20 points) deduct 2 points if missed one homework); Homework on time (10 points, deduct one point for late once)
	Math homework quality		(10 points)Points linked to correct answers
	Chinese homework quality		10 points)Points linked to correct answers
Daily misconduct (90 points maximum)	Late for school		(10 points)- from class register - deduct 1 point per missed day
	Whether left or missed any classes		(10 points) from class register - deduct 1 point per missed class (even negative for more than 10 misses)
	Whether absent from school without legitimate reason		(20 points) from class register - deduct 2 points per absence (even negative for more than 10 absences)
	Quarrel with other students		(40 points) – teacher records and judges
	Disturb others during lessons		(10 points) - teacher records and judges
Daily good behavior (15 bonus points maximum)	Positive good behavior		maximum 3 points - teacher judgment
	Helping teachers		maximum 3 points - teacher judgment
	Helping or caring for classmates		maximum 3 points - teacher judgment
	Good citizenship in general		maximum 3 points - teacher judgment
	Any other good behavior		maximum 3 points - teacher judgment
Total score (130 points plus 15 bonus maximum)			

Note: The maximum is 130 – 40 points for academic results, 90 for bad behavior, and a maximum bonus of 15 for good behavior. Each student's results were discussed between teacher and student in the TSF experiment, and further by parents in the TSPC experiment . This form is completed and discussed bi-weekly during the experiment, a total of 13 times.

Figure 1: Student feedback scorecard

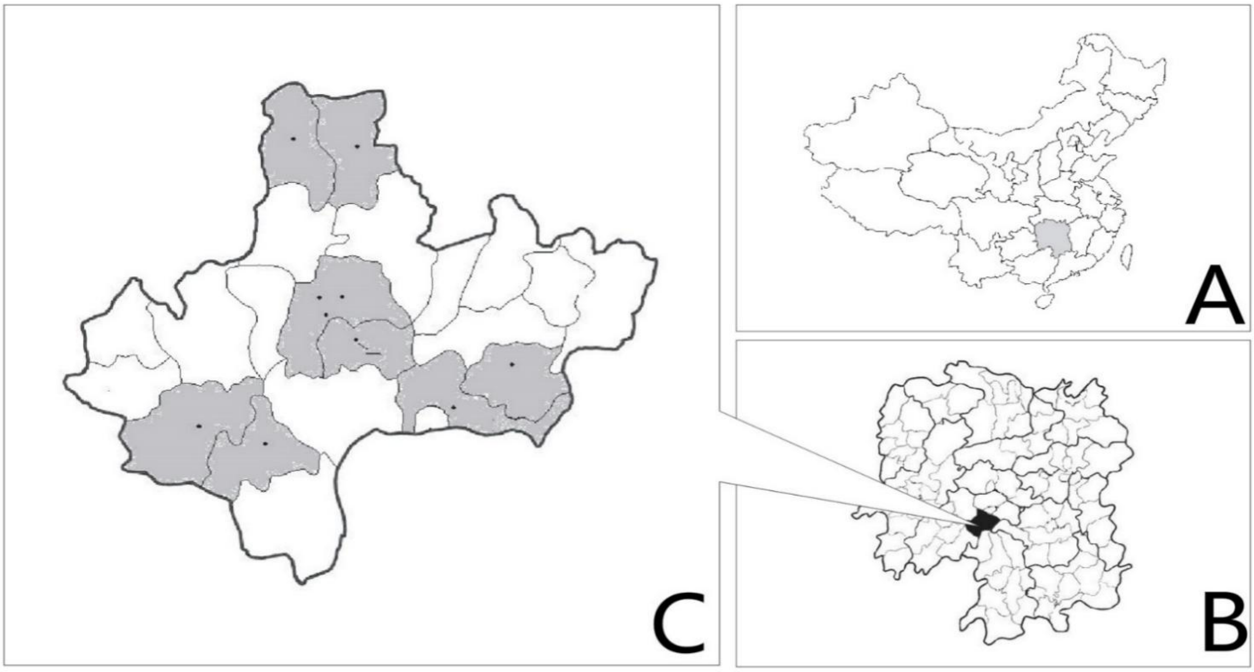

Figure 2: Locations of sample schools in Shaoyang County, Hunan Province, China
Note: Panel A indicates location of Hunan Province in China. Panel B indicates location of Shaoyang County in Hunan province. Panel C indicates locations of sample schools and districts in this study.

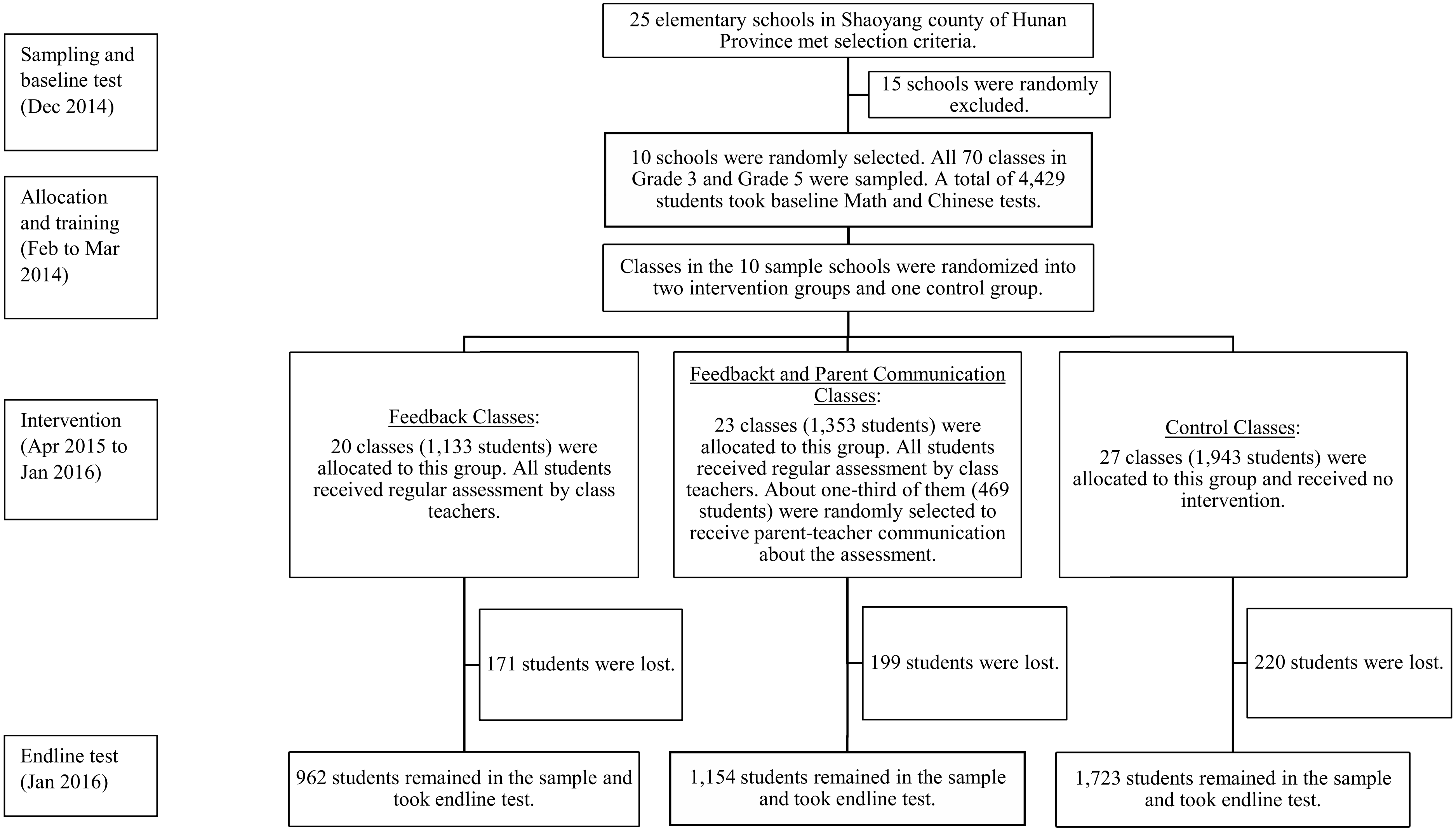

Figure 3. Profile of randomized controlled trials