

Effect of Parental Migration on the Academic Performance of Left-behind Children in Northwestern China

Supplementary Materials 3. Sampling protocol and data collection approach

Sampling

Our sampling frame was restricted to Haidong Prefecture, a poor minority area in Qinghai Province in northwest China. In order to create a sample with enough variation in household migration status to conduct our analysis, we chose to focus our study on poor, rural areas with high population densities and high rates of off-farm employment. A quarter of the population of Qinghai lives in Haidong Prefecture, even though it accounts for only about 2% of the province's total area. Additionally, of the six counties in the prefecture, five of the counties are nationally designated poor counties (NBSC, 2014). For these reasons, Haidong Prefecture was determined to be a suitable location to select our sample.

The next step in the sampling protocol was to choose the sample schools. We obtained a comprehensive list of schools in our six sample counties from each county's education bureau. Based on these lists, we randomly selected 130 schools with classes in grades 1 to 6 in the six sample counties to be included in our sample.

We decided to focus on students in the fourth and fifth grades for two reasons. We believe that students of this age were old enough to be able to fill out their own survey forms and take a standardized examination, but also young enough that they could be followed for a sufficient period of time. In each grade, we randomly selected 2 classes (if there were more than 2 classes in the grade). On average there were 1.3 fourth grade classes and 1.4 fifth grade classes per school. All students in sample classes participated in our survey. In total, the sample included 13,055 students.

Descriptive statistics generated from our data show that the profile of sample students is fairly typical of students from rural areas. Approximately 48.2% of the sample students were girls. In the annual yearbook published by the MOE (2014), girls in rural China account for nearly the same percentage, 47 percent, of each of China's cohorts that are in rural schools.¹ The age of the students ranged between 9 and 18 years in 2013 when we conducted the baseline survey. However, 99% of the students were between the ages of 9 and 13 years.

Although at the time of the baseline survey the sample included a total of 130 schools and 13,055 students, there was some attrition (848 students) by the end of the study, primarily due to school transfers or absence due to illness/injury. This rate of attrition is low compared to other studies conducted with children in rural China (Lai, Luo, Zhang, Huang, & Rozelle, 2015; Mo et al., 2014) and unlikely to impact our findings. By the time of the endline survey in 2014, we were able to follow up with 12,207 students.

Data collection

The research group conducted two waves of surveys in the 130 sample schools. The first round of survey was a baseline survey conducted with all students in all sample schools in September 2013 at the beginning of the academic school year. The second wave of the survey was our endline survey, which was conducted at the end of June in 2014, a time that coincided with the end of the 2013-2014 academic year.

Academic performance

In each wave of the survey, the enumeration team visited all 130 schools and conducted a three-part survey. In the first part students were given a 30-minute standardized English test, the scores of which we used as our measure of student academic performance. Before each round of the survey, we tested the English test

items with over two hundred 4th and 5th grade students to ensure the quality of the baseline and endline English examinations. All the questions in the endline test were different from those in the baseline test. We administered and printed the test ourselves to ensure that it was not possible for the students to prepare for the examination. Also, our enumeration team strictly proctored the test in order to minimize cheating. The team also enforced time limits for the examinations.

We use the standardized English test scores as our measure of academic performance. English test scores were measured during the endline and baseline surveys using a 30 minute English tests. The English tests were constructed by trained psychometricians. Mathematics test items for the endline and baseline tests were first selected from the standardized English curricula for primary school students in China (and Qinghai provinces in particular) and the content validity of these test items was checked by multiple experts. The psychometric properties of the test were then checked using data from extensive pilot testing. We use standardized test scores rather than raw test scores to make student performance comparable across different grades and classes, different periods, and different cohorts. Specifically, in order to standardize each individual observation we subtracted the mean of the comparison group and divided by the standard deviation (SD) of the distribution of the comparison group (the comparison group consists of the households that neither mother nor father out migrated between the two rounds of surveys—for more details of the group, please see the subsection below). Therefore, a standardized score of 0.2 represents someone who scored 0.2 standard deviations above the average of the comparison group. We standardized scores by the grades of the students separately. Appendix Figure 2 depicts the distribution of the standardized baseline English test scores.

We chose English as our subject of study interest for two reasons. First, English is one of the main subjects included as part of the competitive examination system in China that determines entrance into both senior high school and college (Bolton & Graddol, 2012; McKay, 2002). It is a fact that for the past decade or more English takes up nearly the same share as Math and Chinese in China's national high school and college entrance exams (*zhongkao* and *gaokao*). Specifically, the share of English in the overall exam ranges from 20 percent to 25 percent.

Second, English teaching and English learning are particularly weak in poor areas of rural China (G. Hu, 2005; X. Hu, 2009; Li, 2002; H. Zhao, 2003). Studies have shown that a low English score is one of the largest impediments against keeping rural students from attending senior high school in China (Loyalka et al., 2014). Because of this, it must certainly be true that low competency in English would seriously hinder the academic progress of rural students. Due to these reasons we believe that English is an appropriate subject that we can use in our study to measure student academic performance.²

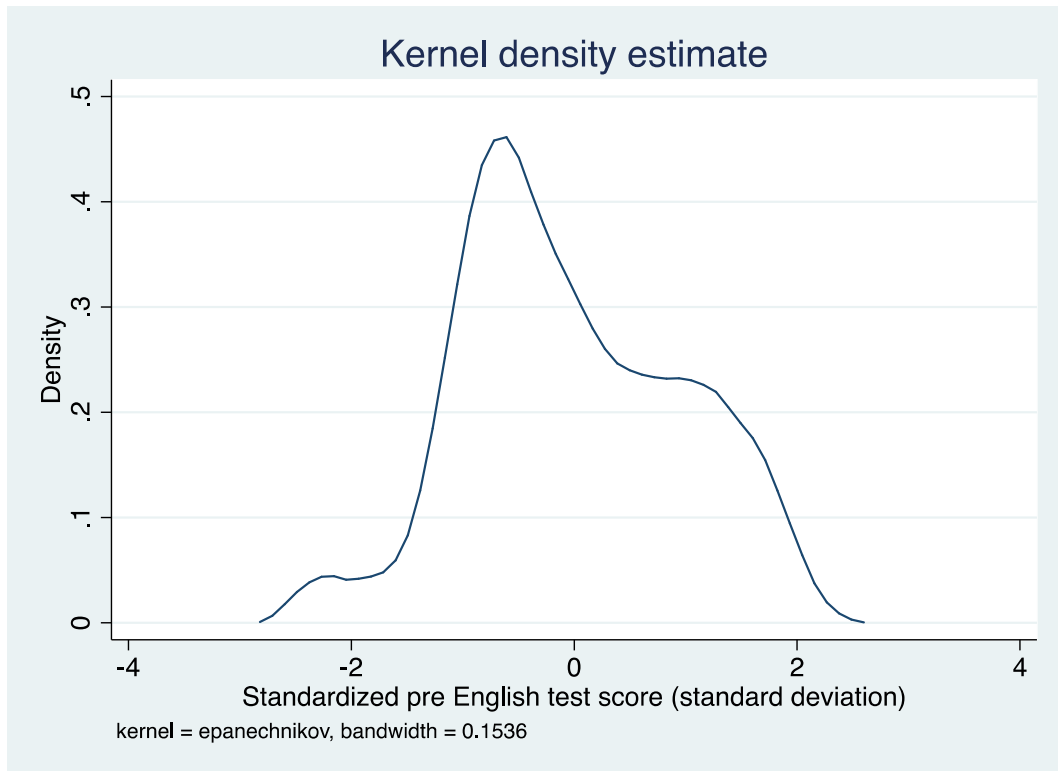
Other covariates

In the third part of the survey we collected data on the characteristics of the sample students. From this part of the survey we were able to create a set of demographic and socioeconomic variables. The dataset includes measures of each student's characteristics, such as *female*, *age*, *ethnic minority*, *5th grade*, *boarding student*. We also created a number of variables measuring family characteristics, including *assets*,³ *father has junior high school or higher degrees*, *mother has junior high school or higher degrees*, and *number of siblings*. This information is beneficial to our research for two reasons: first, it allows us to explore whether the effects of parental migration on the school performance of LBCs are heterogeneous across

children and households; second, these variables may directly affect school performance and by controlling for them we may more efficiently measure the effect of parental migration on school performance.

Notes

1. According to our calculation using data published by statistical yearbooks of Shaanxi, Ningxia, Qinghai, Gansu and Xinjiang, in 2013, girls in rural areas of northwest China also account for nearly the same percentage, namely, 47%, of the class.
2. Before each round of the survey, we tested the English test items with over 200 fourth and fifth grade students to construct baseline and endline English exams. In doing so, our test is with moderate difficulty and high distinction level, as shown in Appendix Figure 2.
3. Asset is calculated by each account of family durable goods multiplying by their prices, then sum all index and take the logarithm.



Appendix Figure 2. Kernal distribution of Standardized baseline English test score.