

Sibling Influence on the Human Capital of the Left-Behind*

Costanza Biavaschi[†] Corrado Giuliatti[‡] Klaus F. Zimmermann[§]

Abstract

While a growing literature has analyzed the effects of parental migration on the educational outcomes of children left behind, this study is the first to highlight the importance of sibling interactions in such a context. Using panel data from the RUMiC Survey, we find that sibling influence on school performance is stronger among left-behind children. Hence, parental migration seems to trigger changes in familial roles and sibling effects among children. However, it is primarily older sisters who exhibit a positive influence on their younger siblings. We corroborate our results by performing a series of tests to mitigate endogeneity issues. The results from the analysis suggest that sibling effects in migrant households might be a mechanism shaping children's outcomes and success and that adjustments within the family left behind have the potential to generate benefits – or reduce hardships – in response to parental migration.

JEL codes: O15, J61

Keywords: Children left behind, siblings, human capital, test scores, Math, Chinese language

*The authors would like to thank participants at the IZA and Center on Human Capital Workshop on Migration and Human Capital (organized in collaboration with the Journal of Human Capital and the IZA Journal of Migration), the 8th IZA-World Bank Conference on Employment and Development, the 5th CIER-IZA Workshop, the 28th ESPE Conference, and seminar participants at Peking University and at the Chinese Academy of Social Sciences. This draft has greatly benefitted from comments by Joseph Altonji, Jinyoung Kim, Peter Kuhn and Yaohui Zhao, by the two editors Isaac Ehrlich and Rodrigo Soares, two anonymous referees, as well as from editorial suggestions by Victoria Finn. The Longitudinal Survey on Rural Urban Migration in China (RUMiC) consists of three parts: the Urban Household Survey, the Rural Household Survey and the Migrant Household Survey. It was initiated by a group of researchers at the Australian National University, the University of Queensland and the Beijing Normal University and was supported by the Institute for the Study of Labor (IZA), which provided the Scientific Use Files. The financial support for RUMiC was obtained from the Australian Research Council, the Australian Agency for International Development (AusAID), the Ford Foundation, IZA and the Chinese Foundation of Social Sciences.

[†]Institute for the Study of Labor (IZA). Biavaschi@iza.org

[‡]Institute for the Study of Labor (IZA). Giuliatti@iza.org

[§]Institute for the Study of Labor (IZA) and University of Bonn. Zimmermann@iza.org

1 Introduction

After decades of research on the factors that induce people to migrate and the consequences for the receiving regions, the literature has now turned its attention to the effects of migration on the sending areas. Within this context, recent studies have focused on the impact of family separation on individuals left behind. A number of papers have analyzed the consequences of a family member's migration on the education outcomes of left-behind children although no consensus has been reached on the sign and magnitude of such an impact. On the one hand, parental emigration inherently implies their absence from the household and can hence have detrimental effects on children's outcomes (Antman Antman). On the other hand, migration usually entails a flow of remittances that might be invested in children's education. Therefore, migration represents a type of parental absence that is fundamentally different from other types of separation, such as death or divorce. Furthermore, the effects of migration on children's outcomes have been shown to be very heterogeneous, varying depending on the age and gender of the child, as well as the gender of the parent (for a review see Antman Antman).

The novel approach of this paper is to study changes in sibship correlation in response to parental migration and, hence, to incorporate the role of sibling interactions in a context where parents are absent due to migration. In practice, we study the effect of a sibling's achievement on another sibling's achievement focusing on how such influences change in migrant versus non-migrant households. Our approach combines two strands of the literature: The first is the migration literature on the left behind, already briefly discussed. The second is the work focusing on the importance of siblings' correlations.

The literature on sibling interactions has found important spillovers in education and income (Solon 1999; Black and Devereux 2011) in several developed countries (Schnitzlein 2014). Sibling correlations capture "nature" and "nurture" effects within households, common environments, as well as the influence of one sibling on another (Black and Devereux 2011). Siblings might affect each other in a number of ways. For instance, sociologists propose that siblings might act as role models (Haynie and McHugh 2003), and psychologists add that they might even directly shape reciprocal outcomes by providing socializing opportunities or directly affecting (younger) siblings' personality and intelligence (Arnold et al. 1975). Economists have shown that older siblings influence younger siblings' high school graduation rates (Oettinger 2000; Rees and Sabia 2009) as well as risk-taking behaviors (Altonji et al. 2010). Furthermore, a long-standing literature shows how birth order, sex composition and family size affect children's outcomes (Behrman et al. 1982; Hanushek 1992; Kaestner 1997).

It is rational to expect that sibling interactions might play a different – if not more important – role during parental absence, particularly in the case of migration. Yet, to the best of our knowledge, the literature studying the impact of such disruptive events on children’s human capital development is scant. Qualitative findings from psychology literature suggest that siblings experience increased closeness as a result of the shared experience of loss of a parental figure (Abbey and Dallos 2004), but very little is known about the interaction between siblings, human capital formation and migration. On the one hand, older siblings might substitute for parents and hence have a stronger influence on younger children when parents are absent. On the other hand, if supervision from parents and older siblings are complements, parental absence will still be detrimental for younger children’s academic achievements.

Our paper explores for the first time the role of sibling spillovers on the educational attainment of left-behind children by focusing on China, a country where parental migration has become a primary concern. Official Chinese statistics report that there were more than 160 million rural migrant workers in the first quarter of 2013 (NBS China 2013). This massive migration has attracted considerable attention regarding its social consequences, specifically because it has created a large number of split households and left-behind individuals. Due to regulations within the hukou – the household registration system – which usually prevent access to quality schools and welfare benefits for non-residents living in urban areas, it is often difficult for migrant parents to bring children along. The All-Women Federation China estimates that around 40 million children 15 and under are left behind in rural areas due to parental migration. This accounts for around 20% of all rural children in China – and corresponds to the total U.S. population of children roughly the same age. Given the size of this phenomenon, it is therefore important to understand the effects of parental absence on children left behind in rural areas. Furthermore, learning whether sibling interactions are an intra-household mechanism that can smooth the potential adverse migration effects on the human capital development of left-behind children is key to understanding the necessity for and scope of public interventions that influence these interactions, e.g., the one-child policy.

This paper, therefore, provides the first evidence of the changing role of sibling interactions in the context of migration. While our results cannot be interpreted as causal estimates (in line with most of the sibling interactions literature), we exploit the richness of our dataset and perform an array of tests to mitigate the concerns over endogeneity. Thus, methodologically we add to the literature in several respects. First, we address the problem of unobserved shared influences using unique panel data on sibling pairs. We are therefore able to control for all biological and environmental time-invariant factors that affect sibling interactions. Second, to address the concerns related to the endogeneity of the migration decision, we

adopt two strategies. We exploit the longitudinal dimension of the data to compare the performance of left-behind and non-migrant children before their parents migrate, carefully controlling for migration reasons. As long as the school performance of children is not driving the migration decision and future migration is not affecting past performance, such a strategy is indicative of the relevance of the endogeneity of migration in influencing our results. We subsequently exploit the differential timing of migration to compare currently left-behind children with those who will be left behind in the future.

Our findings indicate that sibling influence on school performance is stronger among left-behind children, although there is only an effect for Chinese language and not for Math. The results also highlight remarkable heterogeneous patterns depending on the gender composition and the age difference between sibling pairs. Although some left-behind children seem to suffer in terms of school performance, the presence of an older sibling completely balances out such a negative outcome. These results therefore speak to several aspects of the literature. First, parental migration not only results in a change in bargaining power among household decision makers (Antman Antman), but it also triggers changes in the roles and influences between children. Second, peer effects and sibling effects in migrant households are meaningful factors worth studying as they seem to play a particularly important role in shaping children's behavior and outcomes.

The remainder of the paper proceeds as follows. Section 2 briefly outlines the empirical model. Section 3 describes the data used in the analysis. Benchmark results and the heterogeneity analysis are presented in Sections 4 and 5. Section 6 meticulously explores potential econometric issues and outlines our strategies to mitigate endogeneity. Section 7 concludes the paper.

2 Do Siblings Affect Human Capital Accumulation?

The economic literature has long recognized the important role of interactions between parents and children for the child's development. Therefore, parent-child interactions and hence child development are affected by migration as parents are absent and children are often left behind. Consequently, parental migration might trigger changes in resources spent on children's education (Hanson and Woodruff 2003). It could also affect bargaining powers within the household or differentially influence the incentives to invest in children (Antman 2011b). Parents or caretakers might adopt reinforcing or compensating investments in children not only in terms of monetary resources, but also in terms of intra-household task reallocation (Antman 2011a). There is a growing but still scant literature, including studies within the context of developing countries, analyzing the effect of parental migration on children (for a

review, see Antman Antman). For example, using Mexican data, Antman (2012) studies the effect of parental migration on children’s educational attainment employing age of siblings as an instrumental variable and finds positive effects on left-behind daughters. A few studies have explored the outcomes of left-behind children in rural China. Using data from two provinces only, Meyerhoefer and Chen (2011) show that migration is associated with a lag in the educational attainment of girls, which is interpreted as a re-allocation of time spent by girls towards household production within migrant families. Kong and Meng (2010), using cross-sectional data for the year 2008, show that being left behind in rural areas is associated with children’s lower schooling performance (based on the subjective assessment of parents or guardians). Chen et al. (2009) and work by Wang (2012) look at the effect of parental absence on children’s school enrollment, performance and health. While the former study finds no effect on educational achievement in Shaanxi province, the latter finds negative and persistent effects of parental absence on children’s school enrollment, especially for boys. We build upon this work and introduce the effect of sibship on the human capital performance of children left behind in rural China.

While parent-children interaction has been receiving increased attention, the role of sibling interaction in migrant households has not yet been established. The presence of an older sibling might attenuate or reinforce the disruption from migration, hence affecting children’s outcomes, as intra-household adjustments might be borne differently by relatively older or younger children. The paper focuses on this novel mechanism.

In order to empirically test the research question of interest, we estimate regression models with the young children’s school performance (exam scores in Chinese language and Math) as the response variable and with the oldest sibling’s performance, an indicator for being left behind and the interaction between the two latter terms as the key explanatory covariates:

$$Score_{ijt}^Y = \beta_0 + \beta_1 Score_{jt}^O + \beta_2 Left-Behind_{jt} + \gamma (Score_{jt}^O \times Left-Behind_{jt}) + \beta_3 X_{ijt}^O + \beta_4 W_{jt} + \eta_t + c_i + c_{ijt}. \quad (1)$$

We indicate with the superscript Y the outcomes and characteristics of the young children in the household and with the superscript O the score of the oldest child. Hence, $Score_{ijt}^Y$ is the score in Math or Chinese for child i , in household j , in year t . $Left-Behind_{jt}$ is the indicator that equals one when either parent has lived outside the household for at least one month in the 12 months preceding the survey.¹ $Score_{jt}^O$ measures the oldest sibling’s score,

¹In Figure A1 we show the sensitivity of our results to more restrictive definitions (i.e. longer migration

and thus β_1 captures the correlation between the scores of the oldest and younger siblings. The interaction term ($\text{Score}_{jt}^0 \times \text{Left-Behind}_{jt}$) is the key variable of interest, measuring the additional sibling influence if the children are left behind. In our regressions, we cluster standard errors at the household level, accounting for key variables being measured at this stratum.²

The regression model controls for a variety of other factors, captured by the vectors X_{ijt}^0 and W_{jt} . Such factors measure younger children's characteristics such as gender, age dummies, entry grade dummies, school quality and whether they attend boarding school; additionally, they capture family characteristics such as parents' education and labor market characteristics, household characteristics such as household size, land and financial assets, as well as village characteristics and indicators for provinces (for OLS models only). Finally, η_t measures time fixed effects, which have been included as yearly indicators, and c_i captures unobserved individual heterogeneity.

Identifying sibling influence on economic achievement in the migration context is a non-trivial exercise. Causal interpretation of the parameters in equation 1 is prevented by the presence of three empirical issues: (i) there might be several common background characteristics or intra-household allocation of investments that can mediate the siblings' influence on each other, both in migrant and in non-migrant households; (ii) the reflection problem (Manski 1993), i.e. the simultaneous determination of individual outcomes and the threat of reverse causality due to spillover effects from younger to older children; (iii) the decision to migrate is in itself endogenously determined, with individuals self-selecting into that choice and parents possibly being influenced by children school achievement when making the migration decision. Points (i) and (ii) are common issues within the literature examining siblings or peer effects, while point (iii) is related more to the migration literature. In our context instrumental variables that tackle these points are unavailable, but our longitudinal dataset can be exploited to provide indicative evidence on the size of γ and the direction of the possible bias in its estimation.

As a starting point, we estimate equation 1 with Ordinary Least Squares (OLS). However, simple OLS estimation ignores the unobserved individual heterogeneity that might drive young children's performance, capture shared time-invariant biological and environmental factors and be correlated with the probability of being left behind or with the oldest sibling's scores. We therefore estimate the model in equation (1) by a standard fixed effect estimator that purges such heterogeneity through first differences.

spells) of the left-behind indicator.

²As a large share of the families have only two children, we do not expect the clustering at the household level to substantially alter statistical power. We also report the baseline results without clustering in Table A1.

A model that controls for individual heterogeneity seems particularly relevant in this context. Purging c_i partially mitigates problems (i)-(iii). In fact, in our context the vast majority of observations are represented by sibling pairs; hence c_i almost corresponds to a household fixed effect. By first differencing, we are therefore able to control for biological factors and *time-invariant* environmental factors (point (i) above). Additionally, the use of a fixed effects estimator partially reduces the endogeneity of the migration decision stemming from time-invariant family traits that affect educational outcomes as well as the propensity to migrate.

Nonetheless, concerns of reverse causality and of the influence of all time-varying factors might persist. Therefore, caution should be used when interpreting the estimates as causal. With these caveats in mind, we will interpret the parameters as capturing an “equilibrium relationship”. We will provide an array of tests in Section 7 that aim to assess the role of time-varying factors and the importance of reverse causality in our context.

3 Data

Our analysis is based on the RUMiC, a large-scale project conducted in China, comprising the Rural Household Survey (RHS), the Urban Household Survey and a Migrant Household Survey. For our purposes, we extract data from the 2009, 2010 and 2011 waves of the RHS (see Akgüç et al. 2014 for a technical description of the RUMiC panel dataset). The dataset contains detailed information about household members – including those currently migrating out of the village – and comprises socio-demographic characteristics, labor market outcomes, migration history and the family situation prior to leaving the hometown. The data also provide information on educational characteristics and outcomes of children younger than 16 years of age and family members older than 16 who are still in school. We complement these variables with a rich set of controls at the village level, which include the village population, educational expenses, as well as the number of teachers and pupils.

We restrict the sample to children between 5 and 18 years old continuously living in rural areas. In the main analysis, we use pairs of older-younger siblings in which the oldest child in the household is paired with all younger siblings (when there are more than two children in school). To be included in the sample, siblings must be of school age and report scores. The sample is unbalanced, with unbalances mostly attributed to exits from the panel (i.e. children who become older than 18) and to data not being collected in one province in 2011.³

Our outcome variables are young children’s test scores in Chinese and Math in the

³Restricting the analysis to a balanced panel does not alter our conclusions. Table A1 in the Appendix shows the results.

semester prior to the interview and reported by a parent or guardian. Although being self-reported, accuracy of test score reporting is increased by the fact that most Chinese schools give each student a booklet recording scores for each semester during schooling (Chen et al. 2009). Hence, concerns of grade misreporting and recall-bias should be attenuated as such a variable could be interpreted as a record-based variable (Chen et al. 2009). Due to varying scoring standards across provinces and even across schools, we normalize scores by dividing them by the highest obtainable score on the specific test (also reported by the parent or guardian).

The key covariate of interest is an indicator for being left behind. In our baseline regressions, we define a left-behind child as one whose father or mother left the hometown for at least one month during the twelve months preceding the survey. Hence, this variable captures children's exposure to the absence of at least one parent during a period while attending schooling. Each year about one-third of the children have at least one parent who has spent some time outside the household, once again highlighting the relevance of such a phenomenon. The most common destinations for migrants are cities on the East coast of China or in the provincial capitals. Migrants are typically employed in low-skilled occupations mostly within the manufacturing and retail sectors. Persistent hukou regulations determine that most rural-to-urban migrations are temporary. During their migration period, migrants typically only return home for short periods of time, most frequently during the Chinese New Year festivities.

Panel (a) of Figure 1 shows the share of households by the number of children aged 5-18 in rural areas. Of households with children in the relevant age range, about 48% have only one child, while 43% have two children. Larger families are uncommon, with about 7% having 3 children and only about 2% having 4 or 5 children. Panel (b) of Figure 1 shows the share of left-behind children in families with 1, 2, ... 5 kids. Interestingly, the likelihood of migration is similar in households with 1 or 2 children; in both cases, the share of left behind is about 35%. The main estimation sample excludes families with only 1 child and focuses on larger families (2 or more children). Panel (c) of the figure reports the share of non-migrants and those left behind in the sample, as well as the number of periods children were exposed to parental absence. Our sample showed that 35% of children experienced parental absence between 2009 and 2011. About 16% of the sample had either parent away in the previous year, 12% were exposed to two migration spells, while about 6% of the children were continuously exposed to migration. Panel (d) shows with whom children live, differentiating between left-behind children and those from non-migrant families. The majority of left-behind children live with their grand-parents or another relative, while about one-third live with the non-migrant parent.

In Table 1 we show characteristics by migration status, i.e., dividing our sample of young children into those who are left-behind and those from non-migrant families (columns 1 and 2). The share of males in the sample is around 60%, with little difference between the two groups. This high share reflects the consequences of the one-child policy.⁴ The average age of left-behind children is also similar to that of children whose parents have not migrated. There is also rather little variation in the school entry age, averaging around six and a half years old. Around 25%-30% of siblings are in boarding school. There seems to be some variation in the subjective assessment of the children's school quality, with left-behind children less likely to be enrolled in a high-quality school (where the benchmark is the average quality of school in the village).

Table 1 also shows the characteristics of the parents and households in which children live. Mothers and fathers of left-behind children are more likely to engage in wage work. The number of household members and household land size do not vary between the two groups; however, households with children left behind have higher financial assets.

The last rows of the table report the characteristics of the children's village and province distribution. Left-behind children are more likely to come from smaller villages, which spend much less on education and have fewer teachers. The provincial distribution of children left behind differs from that of children of non-migratory parents, reflecting varying migration propensity depending on the geographic location.⁵

The most interesting patterns emerging from the first two columns of Table 1 are the rather small differences in terms of individual observable traits between the left-behind children and those of non-migrants, with the exception of parental characteristics which, however, we will be able to control for through fixed effects estimations. Such small differences reassure the likely absence of a strong selection between migrant and non-migrant households. On the other hand, children in migrant households seem to come from more impoverished villages. In all analyses, we will therefore control for time-varying village characteristics.

The third and fourth columns of Table 1 compare the characteristics of the children in our sample with those of children who are excluded from our analysis. The latter group is largely composed of children from one-child households, but also includes children who have (younger or older) siblings who are currently not in school and hence do not report test scores,

⁴In the majority of rural areas, the policy allows couples to have a second child if their first one is a girl. On aggregate, for families with more than one child, this determines a disproportionate share of males among youngest siblings and females among oldest siblings: In fact, in our sample, nearly 67% of the oldest siblings are female.

⁵For the 2011 wave, data from the Guangdong province were not collected, which slightly unbalances aggregate figures on the geographical distribution. For robustness purposes, we have performed our entire analysis on a restricted sample excluding this province, finding essentially identical results to those presented in the paper.

or siblings whose scores are missing. The purpose of this comparison is to highlight potential selectivity patterns emerging from the definition of our sample. Unlike what was expected a priori, such differences are rather limited. One peculiar aspect is the sex ratio, which again exhibits a relatively large share of males. Within the sample of “Other children,” mostly from one-child families, the reasons for such an unbalance are related to the consequences of sex-selective abortions (for a recent description, see Wei and Zhang 2011). With the exception of village characteristics and provincial distributions, other socio-demographic variables are very similar between the children in our sample and other children.⁶

Table 2 shows the performance in Chinese and Math by left-behind status. For completeness, we also report exam scores of children that fall outside of our estimation sample. Younger siblings perform consistently better than older siblings, with such differences present for both children left behind and those living in non-migrant households. How much of the sibling resemblance is driven by individual heterogeneity, endogeneity of the migration decision or simply observable differences across the two groups? The remainder of the paper seeks to answer these questions.

4 Results

The key results of the paper are reported in Tables 3 and 4. Starting with the OLS estimates, some interesting patterns arise. Looking at the performance in Chinese language throughout all specifications, there is a high correlation between the oldest and the younger child’s score, with estimates in the range of 0.39 to 0.45. This estimated correlation is remarkably in line with sibling correlations found in other studies (see, e.g., Björklund and Salvanes 2010). As shown in columns II and III of Table 3, on average the parental absence does not seem to have a large effect on children’s performance in school, which might be driven by remittances balancing the potential adverse effect of being left behind. Columns IV and V show the estimates of the interaction between the older sibling’s scores and being left behind with and without controls for several characteristics of the child, parents, household and village. Both the indicator for being left behind and the interaction term are now sizeable in magnitude and statistically significant.

To take a practical example, let us consider a left-behind child and a child whose parents have not migrated, both of whom have an older sibling who scored zero in Chinese (this example purely serves the scope of comparative statics, as there are virtually no children exhibiting a score of zero). Compared to children of non-migrants, young left-behind children

⁶For example, the one-child policy was particularly enforced in rural areas of Jiangsu, Chongqing and Sichuan. This determines that one-child families are relatively more represented in these provinces.

with a poor performing oldest siblings face a score penalty of 12% on average. However, as younger children are exposed to better performing siblings, not only are their scores positively correlated to their sibling's scores, but also the disadvantage of being left behind is fully compensated. Therefore, left-behind children with average performing siblings end up suffering from only a 1%-1.5% penalty in their scores. Thus, the presence of a top-performing sibling fully re-balances the score disadvantage stemming from being left behind.

In conclusion, sibling influence is stronger in left-behind households since sibling correlations are higher than in non-migrant households. All other controls exhibit the expected sign: Children enrolled in high-quality schools have higher scores; furthermore, the higher the number of teachers in the village, the better the children perform. Two variables inhibit children's scores: a high number of students in the village and the household's land size. Villages with too many students might have lower quality schools and education systems less aligned with the children's specific needs. Children from households owning more land might face a disincentive to invest in education, and/or might be induced to work (Liang and Chen 2007).

Columns V to VIII of Table 3 also show the same regressions for performance in Math, with estimates for the sibling correlation very similar to those for Chinese language. However, the estimates for the left-behind indicator and for its interaction with the oldest sibling's score are smaller in magnitude and statistically insignificant despite following a similar pattern in terms of sign. These results are not entirely surprising if one considers the different skills – and their transferability – involved in the study of Chinese and Math. We could postulate that good performance in Math is harder to achieve and can hardly be influenced by peers, while Chinese performance – driven by a student's ability to read and write a certain number of characters – can be better monitored and influenced by older siblings. The result that Math performance is more prone to the influence of factors other than sibship is broadly compatible with evidence by Zimmerman (2003), who finds that peer effects are stronger with verbal scores than with Math scores; by Metzler and Woessmann (2012), who find a positive effect of teacher quality on children's Math performance but not on reading; and by Angrist et al. (2012), who find that a schooling program had a larger impact on Math achievements than on reading. The different patterns between Chinese and Math scores, and the statistical insignificance of the latter, persist throughout most of our analyses.

As previously highlighted, these patterns might suffer from the presence of individual effects that correlate with the covariates of interest and which could stem from individual ability as well as from the endogeneity of the migration decision. In Table 4, we tackle this issue by estimating fixed effect regressions. The fixed effect estimator purges out any observed and unobserved variation that is common within individuals. In our context, where the

majority of sibling pairs are from a two-sibling household, fixed effect results are also almost identical to results obtained by including household time-invariant effects.⁷ Hence, we are able to control for any individual- or household-specific propensity to migrate. Remarkably, the fixed effects estimates reinforce our previous findings.

As in the OLS case, left-behind children seem to suffer from parental absence in their performance in Chinese, while there are little effects in Math. The sibling correlation remains sizeable even after controlling for individual heterogeneity. The introduction of the fixed effects exacerbate the older sibling's influence on Chinese scores when parents are away: The estimate for the interaction term between the left-behind indicator and the oldest sibling's score doubles compared to the OLS estimates. Achievement in Chinese of the left-behind children is around 30% lower than that of the non-migrant children in the most conservative specification, independent of the controls added. However, as before, left-behind children seem to resemble their older siblings in such a way that this disadvantage is reduced and brought to zero in the case of top-performing older siblings. As before, performance in Math is unaffected by being left behind, and sibling influence is the same in both the migrant and non-migrant households.

Our baseline analysis suggests two main findings. First, the presence of an older child plays a more important role in terms of school performance when the parent has migrated. Second, there is heterogeneity between school subjects in terms of the role that both parental absence and sibling influence play in shaping the human capital development of the left-behind.

5 Heterogeneity Analysis

We continue our analysis showing the heterogeneity of the results by children and parent characteristics in Table 5. The first panel in Table 5 shows performance by the sex composition of sibling pairs. In both Chinese and Math, older male siblings do not exhibit additional effects on the left-behind. However, the opposite is true for older female siblings. In this case, the positive correlation across scores is amplified among left-behind households. Once again, this pattern is strong in the case of Chinese performance. However, when focusing on female-female pairs, we also find sibling effects in the case of Math performance, albeit these results are statistically significant at the 10% level only. Therefore, it appears that older sisters exhibit a nurturing effect – large enough to even reverse the penalty of being left behind – on younger siblings, who seem to have an advantage from the presence of a top-performing child. The changing role of female children in the household quantifies the

⁷Results available upon request.

speculative conclusion in Meyerhoefer and Chen (2011) who suggest that migrant households re-allocate girls' time towards household production. Additionally, these results are also in line with evidence showing that older sisters share significant child-care responsibilities in many developing countries (Levison and Moe 1998; Ilahi 2001; Edmonds 2006), hence potentially greatly influencing their siblings' development. Lastly, our results are also in line with those in Qureshi (2011) who shows that schooling of older sisters has beneficial impacts on the education of their younger brothers.

We further disentangle the effects by looking at which stage of compulsory education – in earlier or later grades – sibling influence matters the most and at which age differential such effects come into play. The second panel of Table 5 presents results for children with a smaller (equal or less than 5 years) and larger (more than 5 years) age distance between the sibling pairs. As one would perhaps expect, the sibling influence is stronger among children whose age differential is further apart. Correlations are particularly stronger in the left-behind group in both Chinese and Math. For children closer in age, correlations are not only statistically insignificant but also small in terms of magnitude. With this result also holding in the acquisition of Math skills, we could conjecture once again that this pattern captures a nurturing effect of older siblings on younger ones.

The last four columns of this panel show the effects for younger children in grades 1 to 3 and grades above 3. Remarkably, children who have already spent a few years in school are most affected by sibling influence. While results are non-significant for children in grades 1 to 3, children in grades 4 and above exhibit: i) a strong correlation across Chinese exam scores, ii) a penalty if left behind, and iii) stronger dependence on older siblings' performance if left behind – although these effects are weaker than those by age. As before, the left-behind disadvantage is fully compensated for by the average-performing older sibling. Overall, it seems that the effect on younger siblings is driven by both differences in age and the younger child's stage on the educational ladder.

In the third panel of Table 5, we show whether the impact is stronger in households where the child lives with parents or other individuals (relatives, friends or other people in the village). As expected, the impact of migration and the sibling effects are stronger when parental authority is absent and children live with members outside the restricted familial nucleus.⁸ We subsequently check whether sibling correlations are stronger in households with a previous migration experience. We select a subsample of households with a migration

⁸While this paper highlights the changing role of siblings in response to parental migration, there could also be readjustments across other left-behind household members. An additional venue for further research could explore, for instance, how grandparents or other family members influence children's development (along the lines of our heterogeneity tests in this section). Studying this channel might bring additional evidence regarding the effect of migration on families left behind.

history (i.e. those households who report having migrated before 2009) and those without. Remarkably, only children whose parents had not previously migrated are those who face a penalty for being left behind. As before, however, this penalty is mitigated by the presence of siblings. Therefore, it appears that the intra-household reallocation of roles (and potentially resources) is only temporary, and migrant households are able to adjust over time to the possible disruption of migration.

In the last panel of Table 5, we test whether sibling effects depend on which parent is absent. We show estimates from models in which we consider left behind children only those whose mother is absent (columns I and IV), the father is absent (columns II and IV) and both parents are absent (columns III and VI). Remarkably, we find similar effects on the left behind independently of which parent migrates, although the mother matters somewhat more for Chinese than the father. Once again, the impact is essentially zero in the case of Math performance.

6 Econometric Issues

As mentioned in Section 2, there are three main methodological challenges in our analysis. First, the presence of omitted *time-varying* environmental factors between siblings might affect the results. Second, it might be argued that reverse causality is present with older and younger children simultaneously influencing each other. Third, the migration decision might remain endogenous even after controlling for several confounding factors, as we do in our regressions. Migrants self-select in the decision to migrate, and there might be additional unobserved factors correlated with both children's outcomes and migration status. For instance, families with better socioeconomic status, networks and earning potential might be more (or less) likely to migrate as well as possibly provide an environment in which siblings might interact more (or less) with each other. Furthermore, reverse causality might also be present in this decision, with migrants moving and leaving children behind due to children's school performance. For example, parents might want to provide better educational opportunities for children or might decide not to move if children are facing particular difficulties in school. Although we control for a significant variety of characteristics, it is difficult to rule out a priori that endogeneity affects the results.

In this section, we proceed with a few robustness checks which we present in Table 6. We aim to detect the relevance of these problems in the context of our study. Before proceeding, it should be noted that one of the main advantages of our dataset is the panel structure, which allows us to observe the children at three points in time. Hence, we have already purged from the model time-invariant heterogeneity in both the migration decision and the

siblings' relationship by using the fixed effect estimator. The OLS estimator appeared to be upwardly biased (as we would expect if siblings had a positive influence on each other), although results from the OLS and fixed effect models are not very different in terms of the magnitude of the estimated coefficients. Albeit not a test, our baseline results suggest that reverse causality and endogeneity in migration might not be fundamental concerns given the controls and structure of the data.

Exploiting variation in the timing of migration to mitigate endogeneity. Children in our panel are observed as being left behind over different years. Instead of using variation from comparing left-behind and non-migrant households, we use variation in the timing of being left behind and drop observations of those who never migrate. In this restricted sample, the control group consists of children who are not left behind at time t but were either left behind in the past or will be in the future. Focusing on the left-behind for some t is informative about the presence of the selection between stayers and movers. It should be noted, however, that this strategy will not control for dynamic selection concerns.⁹ In fact, once controlling for relevant characteristics, the key identifying assumption is the absence of selection that changes over time. The first four columns in the upper panel of Table 6 show the results after applying both OLS and fixed-effect estimators, which are similar to those found in the main part of the paper. This particular exercise provides suggestive evidence that results are not driven by the static selection into migration; however, the test remains silent on the likelihood of dynamic selection.

Prospective migration to understand selection. As a secondary check to assess the magnitude of self-selection in migration, we estimate the following model:

$$Score_{ijt}^Y = \beta_0 + \beta_1 Score_{jt}^O + \beta_2 Left-Behind_{jt+1} + \gamma (Score_{jt}^O \times Left-Behind_{jt+1}) + \beta_3 X_{ijt}^Y + \beta_4 W_{jt} + c_{ijt}. \quad (2)$$

An established literature explores how migrants are selected by comparing the pre-migration outcomes of future migrants with those of stayers (Moraga 2011; Kaestner and Malamud 2014). The model above proceeds in the same manner: Assuming that future migration is unrelated to current shocks, and if selection was not a serious concern in our data, we should expect outcomes not to differ depending on migration status. The last two columns of the upper panel show that migration is statistically unrelated to current performance, and hence pre-migration outcomes between migrants and stayers are not different. Furthermore, this

⁹Other studies have used such a technique; see for example Bertrand and Mullainathan (1999).

analysis also suggests that parents do not respond to their children's achievement in school by adjusting their migration plans.

Understanding reverse causality in migration. We further investigate this last point by directly testing whether migration depends on educational success. We relate the probability of being left behind at time t with children's educational achievement, controlling for the same characteristics used in the baseline specification. We therefore correlate the probability of parents moving in response to their children's current educational outcomes. Following the approach adopted throughout, we estimate the model through fixed effects.

Columns I and II in the lower panel of Table 6 show the estimates from such a model. Current performance in both Chinese and Math is unrelated to the probability of staying behind, not only statistically but also in terms of magnitude of the coefficients. As an additional test, we perform the same analysis, but relate the likelihood of being left behind at $t+1$ with current performance. In other words, we look at whether children's education performance is related with the future migration decision. Results are reported in columns III and IV and show that children's current performance in Chinese and Math are also unrelated to future parental migration. Therefore, concerns of reverse causality should not be strong.

Understanding the importance of the reflection problem. Lastly, as already mentioned, younger siblings' performance might in turn affect older siblings' performance. Although this paper does not aim to pinpoint the "directionality" of sibling interactions, we check whether the oldest sibling responds to his or her changing role in the household. In this test, the outcome variable is the score of the oldest sibling, while the key covariate is an interaction term between the left-behind indicator and the average score of the younger siblings. The last two columns of Table 6 show that the impact on the older sibling is statistically zero and very small. Although it might come as a surprise, this finding is in line with causal results in the literature on sibling correlation in education both in developed (Oettinger 2000) and developing countries (Nicoletti and Rabe 2014).

Further robustness checks. We have performed a series of additional robustness checks in order to further corroborate our finding. These additional results are presented in the Appendix. In Table A1 we explore several alternative specifications related to different subsamples and specifications. In the top right panel we perform our regression on the balanced sample, i.e., when each child is observed for three consecutive waves. Next to this, we look at a subsample of households with only two children reporting test scores, i.e., we exclude

households with three or more children in education. While point estimates differ from the benchmark model, the pattern of the results is essentially unchanged. In the second panel of Table A1 we first experiment with a specification without clustering the standard errors at the household level, and we then estimate a full interaction model in which the left-behind indicator is interacted with both time-varying and time-invariant individual, household and village characteristics. Both specifications exhibit a pattern remarkably similar to the benchmark model. In the bottom panel, we split our sample by the level of the test score observed in the first wave. The rationale is to ascertain whether ability, here proxied by initial test scores, is a driving force for the decision of the parents to migrate. If this were the case, estimates would be substantially different across the sample of children with higher and lower initial test scores. While we find that the size of the Math estimates is slightly different between the two samples, the estimates for the performance in Chinese, especially in the fixed effects model, are identical.

We have also explored the sensitivity of our results to the definition of the left-behind variable. In our benchmark model, the left-behind indicator measures parental migrations of at least one month, encompassing both short- and long-term absences. Since our data contain the monthly duration of the migration spell, we experiment with all possible definitions of being left behind from the shortest possible absence (at least one month, our benchmark) to the longest one (twelve months). Figure A1 shows the results of this exercise. The upper values of each graph correspond to the estimates of γ , the key parameter of the interaction term between the left-behind indicator and the test scores, while the lower values refer to estimates of β_2 , the parameter pertaining to the left-behind indicator. The graph unequivocally shows that the estimates are stable – and within the same confidence interval – throughout the various definitions. The only exceptions are the estimates pertaining to the most restrictive definition of migration, i.e. 12 months. Since only 2% of the children in the sample (equivalent to 36 observations) fall under this definition, it is not surprising that both the left-behind dummy and the interaction term are estimated with large noise.

Finally, in Tables A2 and A3 we re-estimate, respectively, the specifications of Table 3 and 4 by progressively adding controls in order to check the sensitivity of the estimates to observable factors. Once again, the pattern of the estimates is remarkably stable and similar to the benchmark model.

7 Conclusions

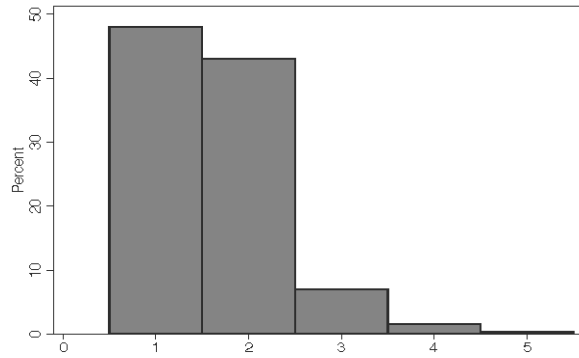
While a growing literature has analyzed the effects of parental migration on the educational outcomes of left-behind children, this study is the first to highlight the importance of sibling

influence in such a context. We find consistent results showing that sibling interactions affect the cognitive development of younger children. The effects are stronger among left-behind children, albeit only in the acquisition of language ability, in which case the positive influence of older siblings compensates for the negative effects of being left behind. Parental migration, hence absence, seems to trigger changes in the reciprocal roles and interactions between children. Heterogeneity analysis reveals that such changes primarily involve the role of older sisters and are stronger depending on the current grade and the age difference between siblings. Additionally, effects are somewhat stronger when the mother migrates instead of the father.

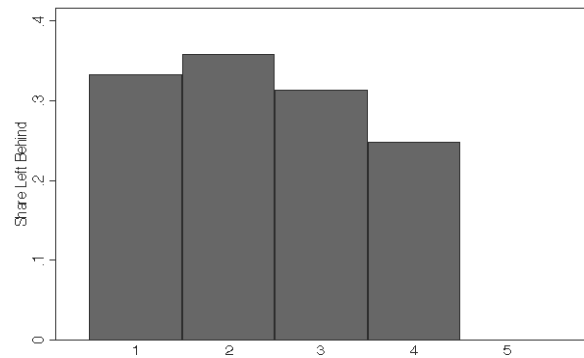
Our results suggest that sibling effects in migrant households play an important function in shaping children's educational outcomes and success. There are several relevant policy implications to consider. With internal migration expected to increase, the welfare of left-behind individuals will become increasingly central to the Chinese government's policy agenda. Our results suggest that policies fostering sibling interactions will sort – *ceteris paribus* – positive effects in terms of human capital development. The relaxation of the one-child policy announced by the Chinese government in November 2013 has the potential to create such externalities. Allowing a larger number of rural families to have more than one child may create the grounds for cultivating sibling interactions.

Figures and Tables

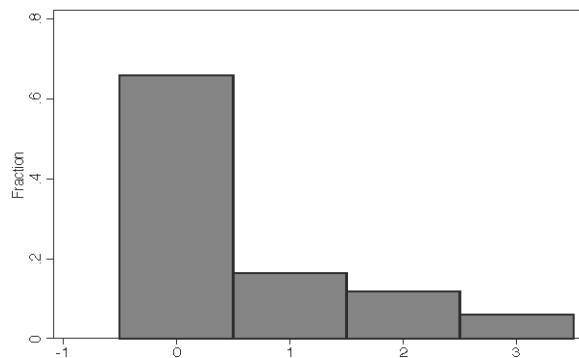
Figure 1: Characteristics of Children in the Household, age 5-18



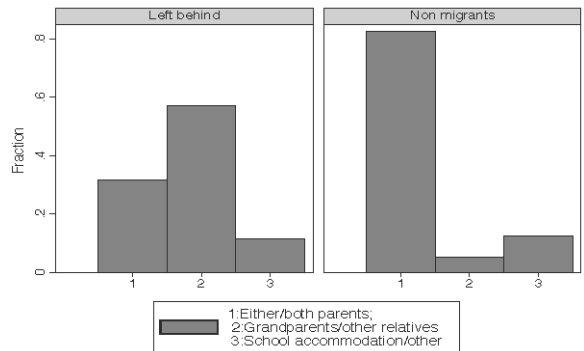
(a) Number of Children in the Household



(b) Fraction Left Behind, by Number of Children in the Household



(c) Fraction of Never Left Behind, Left Behind for 1, 2 or 3 periods



(d) Children Living with Parents/Grandparents by Migration Status

Table 1: Characteristics of Left-Behind Children and Children in Non-Migrant Households

Variable	Young Siblings (in sample)			Other Children (not in sample)
	Left-behind	Non-migrant	All	
Male (D)	0.597 (0.491)	0.618 (0.486)	0.611 (0.488)	0.617 (0.486)
Age	11.041 (3.072)	11.370 (3.104)	11.258 (3.097)	11.721 (3.724)
Age at entry	6.560 (0.735)	6.746 (0.803)	6.683 (0.785)	6.658 (1.293)
Boarding school (D)	0.253 (0.435)	0.298 (0.458)	0.283 (0.451)	0.371 (0.483)
High quality school (D)	0.191 (0.394)	0.251 (0.434)	0.231 (0.421)	0.298 (0.457)
Grade	4.968 (2.730)	5.269 (2.997)	5.167 (2.912)	6.378 (3.583)
Father is farmer or out of labor force (D)	0.103 (0.304)	0.444 (0.497)	0.328 (0.470)	0.326 (0.469)
Father is employee (D)	0.758 (0.428)	0.375 (0.484)	0.506 (0.500)	0.551 (0.498)
Father is self-employed (D)	0.139 (0.346)	0.181 (0.385)	0.167 (0.373)	0.123 (0.329)
Mother is farmer or out of labor force (D)	0.489 (0.500)	0.726 (0.446)	0.645 (0.479)	0.549 (0.498)
Mother is employee (D)	0.429 (0.495)	0.201 (0.401)	0.278 (0.448)	0.387 (0.487)
Mother is self-employed (D)	0.082 (0.275)	0.074 (0.261)	0.077 (0.266)	0.064 (0.245)
Household size	5.155 (1.278)	4.997 (1.183)	5.051 (1.218)	4.304 (1.264)
Household financial asset (ln RMB)	6.983 (2.813)	7.126 (3.356)	7.077 (3.181)	6.809 (3.481)
Household land size (Mu/1000)	0.160 (0.276)	0.147 (0.303)	0.152 (0.294)	0.166 (0.339)
Lag village population (/100)	2.605 (2.009)	3.156 (3.043)	2.968 (2.746)	2.496 (1.673)
Lag village expenditure on education (Mill RMB)	0.005 (0.034)	0.016 (0.131)	0.012 (0.109)	0.010 (0.077)
Lag village number of teachers (/100)	0.109 (0.153)	0.240 (0.921)	0.195 (0.755)	0.095 (0.266)
Lag village number of students (/100)	2.590 (3.880)	3.651 (6.256)	3.289 (5.583)	1.786 (3.736)
Province: Hebei	0.026 (0.160)	0.072 (0.258)	0.056 (0.230)	0.057 (0.232)
Province: Jiangsu	0.111 (0.314)	0.080 (0.271)	0.090 (0.286)	0.167 (0.373)
Province: Zhejiang	0.021 (0.142)	0.065 (0.247)	0.050 (0.218)	0.129 (0.336)
Province: Anhui	0.212 (0.409)	0.155 (0.362)	0.174 (0.380)	0.120 (0.325)
Province: Henan	0.131 (0.338)	0.135 (0.342)	0.134 (0.340)	0.105 (0.306)
Province: Hubei	0.056 (0.231)	0.067 (0.250)	0.063 (0.243)	0.121 (0.327)
Province: Guangdong	0.200 (0.401)	0.349 (0.477)	0.298 (0.458)	0.064 (0.245)
Province: Chongqing	0.026 (0.160)	0.023 (0.151)	0.024 (0.154)	0.064 (0.245)
Province: Sichuan	0.217 (0.413)	0.055 (0.229)	0.111 (0.314)	0.173 (0.378)
Observations	534	1032	1566	4088

Source: RUMiC data, RHS waves 2009, 2010 and 2011. Standard deviations shown in parentheses. Left-behind children are those whom had either parent leave the hometown for at least one month during the twelve months preceding the survey. "Other Children" refers to the group of children not included in the analysis. See text for details.

Table 2: Test Scores in Chinese language and Math by Migration Status

	Chinese		Math	
	Left-Behind	Non-Migrant	Left-Behind	Non-Migrant
Young sibling's score	0.802 (0.119)	0.811 (0.110)	0.818 (0.122)	0.821 (0.116)
Oldest siblings' score	0.776 (0.120)	0.781 (0.119)	0.790 (0.127)	0.795 (0.132)
Observations	534	1032	534	1032
Other children's score (not in sample)	0.822 (0.120)	0.816 (0.122)	0.837 (0.119)	0.833 (0.125)
Observations	1359	2729	1359	2729

Source: RUMiC data, RHS waves 2009, 2010 and 2011. Standard deviations shown in parentheses. Scores refer to exam scores in Chinese language and Math in the semester prior to the interview, as reported by the child's parent or guardian.

Table 3: Performance in Chinese and Math, OLS Results

	Chinese					Math				
	I	II	III	IV	V	VI	VII	VIII	IX	X
Score _{jt} ^O	0.454*** (0.036)		0.453*** (0.035)	0.406*** (0.040)	0.393*** (0.039)	0.410*** (0.028)		0.409*** (0.028)	0.392*** (0.032)	0.378*** (0.032)
Left-Behind _{jt} (D)		-0.014** (0.007)	-0.010* (0.006)	-0.113* (0.058)	-0.120** (0.059)		-0.008 (0.007)	-0.004 (0.006)	-0.045 (0.051)	-0.059 (0.052)
Left-Behind _{jt} × Score _{jt} ^O				0.132* (0.071)	0.140* (0.072)				0.053 (0.060)	0.072 (0.061)
Male (D)	-0.005 (0.005)	-0.007 (0.006)	-0.005 (0.005)	-0.006 (0.005)	-0.008 (0.005)	0.000 (0.006)	0.001 (0.006)	0.000 (0.006)	0.000 (0.006)	-0.001 (0.006)
Age	-0.003* (0.002)	-0.005** (0.002)	-0.003* (0.002)	-0.003* (0.002)	-0.004** (0.002)	-0.002 (0.002)	-0.005** (0.002)	-0.002 (0.002)	-0.002 (0.002)	-0.003* (0.002)
Age at entry	-0.006 (0.004)	-0.012** (0.005)	-0.007* (0.004)	-0.008* (0.004)	-0.010** (0.004)	-0.009** (0.004)	-0.015*** (0.005)	-0.009** (0.004)	-0.009** (0.004)	-0.011*** (0.004)
Boarding school	-0.007 (0.006)	-0.010 (0.007)	-0.007 (0.006)	-0.007 (0.006)	-0.006 (0.006)	-0.008 (0.007)	-0.007 (0.008)	-0.008 (0.007)	-0.008 (0.007)	-0.004 (0.007)
High quality school	0.034*** (0.006)	0.048*** (0.007)	0.034*** (0.006)	0.034*** (0.006)	0.028*** (0.006)	0.037*** (0.007)	0.052*** (0.007)	0.037*** (0.007)	0.037*** (0.007)	0.031*** (0.006)
Grade	-0.002 (0.002)	-0.002 (0.002)	-0.002 (0.002)	-0.002 (0.002)	-0.001 (0.002)	-0.003 (0.002)	-0.002 (0.002)	-0.003 (0.002)	-0.003 (0.002)	-0.002 (0.002)
Father is employee					0.002 (0.006)					0.002 (0.007)
Father is self-employed					-0.004 (0.008)					0.000 (0.009)
Mother is employee					0.012* (0.007)					0.009 (0.007)
Mother is self-employed					0.009 (0.012)					0.012 (0.014)
Household size					-0.003 (0.003)					-0.005** (0.003)
Household financial asset					-0.001 (0.001)					-0.001 (0.001)
Household land size					-0.028** (0.012)					-0.030** (0.012)
Lag village population					0.000 (0.002)					0.001 (0.002)
Lag village expenditure on education					0.005 (0.021)					-0.018 (0.020)
Lag village number of teachers					0.015*** (0.004)					0.013** (0.006)
Lag village number of students					-0.002** (0.001)					-0.001 (0.001)
R ²	0.31	0.10	0.31	0.32	0.35	0.29	0.09	0.29	0.29	0.32
N	1566	1566	1566	1566	1566	1566	1566	1566	1566	1566

Source: RUMiC data, RHS waves 2009, 2010 and 2011. Robust standard errors clustered at the household level in parentheses.
 Column IV, V, IX and X include province fixed effects.
 All regressions include year fixed effects.

Table 4: Performance in Chinese and Math, Fixed Effects Results

	Chinese					Math				
	I	II	III	IV	V	VI	VII	VIII	IX	X
Score _{jt} ^o	0.432*** (0.071)		0.431*** (0.071)	0.331*** (0.075)	0.331*** (0.072)	0.403*** (0.042)		0.402*** (0.042)	0.364*** (0.053)	0.363*** (0.053)
Left-Behind _{jt} (D)		-0.018 (0.013)	-0.013 (0.011)	-0.208** (0.090)	-0.219** (0.090)		-0.009 (0.016)	-0.004 (0.013)	-0.086 (0.068)	-0.093 (0.067)
Left-Behind _{jt,x} × Score _{jt} ^o				0.249** (0.115)	0.256** (0.114)				0.103 (0.081)	0.106 (0.079)
Boarding school	0.009 (0.009)	0.002 (0.011)	0.009 (0.009)	0.008 (0.009)	0.006 (0.009)	0.004 (0.010)	0.001 (0.012)	0.003 (0.011)	0.004 (0.011)	0.002 (0.011)
High quality school	0.024** (0.011)	0.026** (0.012)	0.024** (0.011)	0.023** (0.011)	0.021** (0.011)	0.021* (0.012)	0.023* (0.013)	0.021* (0.012)	0.020 (0.012)	0.021* (0.012)
Grade	-0.003 (0.002)	-0.000 (0.003)	-0.003 (0.002)	-0.002 (0.002)	-0.002 (0.002)	-0.005* (0.003)	-0.004 (0.004)	-0.005* (0.003)	-0.005 (0.003)	-0.005 (0.003)
Father is employee					0.013 (0.014)					0.016 (0.014)
Father is self-employed					-0.002 (0.019)					0.021 (0.020)
Mother is employee					0.009 (0.015)					-0.002 (0.014)
Mother is self-employed					-0.022 (0.025)					-0.027 (0.024)
Household size					0.003 (0.011)					-0.007 (0.013)
Household financial asset					-0.002 (0.002)					-0.000 (0.002)
Household land size					-0.015 (0.014)					-0.038** (0.016)
Lag village population					0.003 (0.007)					-0.008 (0.007)
Lag village expenditure on education					0.078 (0.088)					0.026 (0.057)
Lag village number of teachers					0.021 (0.017)					0.021 (0.017)
Lag village number of students					-0.002 (0.001)					0.001 (0.001)
R ²	0.25	0.02	0.25	0.27	0.29	0.21	0.02	0.21	0.22	0.23
N	1566	1566	1566	1566	1566	1566	1566	1566	1566	1566

Source: RUMiC data, RHS waves 2009, 2010 and 2011. Robust standard errors clustered at the household level in parentheses. All regressions include year fixed effects.

Table 5: Heterogeneity

	By Sex of Sibling Pairs							
	Chinese				Math			
	MO-MY	MO-FY	FO-MY	FO-FY	MO-MY	MO-FY	FO-MY	FO-FY
Score _{jt} ^o	0.2317** (0.0974)	0.5187*** (0.0609)	0.2803** (0.1390)	0.2467* (0.1333)	0.2989*** (0.1020)	0.4409*** (0.0584)	0.4409*** (0.0928)	0.1945 (0.1242)
Left-Behind _{jt} (D)	-0.1564 (0.1011)	0.0882 (0.0979)	-0.3113** (0.1569)	-0.3318** (0.1508)	0.0076 (0.1151)	0.0761 (0.1678)	-0.0911 (0.1220)	-0.3268** (0.1549)
Left-Behind _{jt} × Score _{jt} ^o	0.1948 (0.1463)	-0.1005 (0.1178)	0.3918* (0.2016)	0.3657** (0.1809)	0.0191 (0.1455)	-0.0642 (0.2021)	0.1113 (0.1449)	0.3538* (0.1804)
N	264	248	693	361	264	248	693	361
	By Age Distance				By Grade of Younger Siblings			
	Chinese		Math		Chinese		Math	
	≤ 5 years	> 5 years	≤ 5 years	> 5 years	Grade ≤ 3	Grade > 3	Grade ≤ 3	Grade > 3
Score _{jt} ^o	0.3584*** (0.0978)	0.1688** (0.0681)	0.4221*** (0.0638)	0.1106 (0.0784)	0.2563*** (0.0978)	0.3632*** (0.0938)	0.2353* (0.1311)	0.3591*** (0.0617)
Left-Behind _{jt} (D)	-0.1812** (0.0902)	-0.4149** (0.1938)	-0.0562 (0.0743)	-0.3133** (0.1328)	-0.2081* (0.1202)	-0.2206* (0.1198)	-0.1256 (0.1344)	-0.1345* (0.0731)
Left-Behind _{jt} × Score _{jt} ^o	0.2005* (0.1151)	0.5268** (0.2448)	0.0687 (0.0924)	0.3590** (0.1480)	0.2314 (0.1547)	0.2768* (0.1504)	0.1368 (0.1647)	0.1525 (0.0927)
N	1053	513	1053	513	498	1068	498	1068
	Lives with Parents or Others				Migrated Before			
	Chinese		Math		Chinese		Math	
	Parents	Others	Parents	Others	Yes	No	Yes	No
Score _{jt} ^o	0.2576*** (0.0809)	0.3097 (0.2236)	0.3526*** (0.0591)	0.1786 (0.1399)	0.3397*** (0.0822)	0.0918 (0.1540)	0.3735*** (0.0625)	0.2851*** (0.0741)
Left-Behind _{jt} (D)	-0.0588 (0.1034)	-0.3513* (0.1909)	0.0850 (0.1626)	-0.3111** (0.1324)	-0.0983 (0.0840)	-0.5643*** (0.1607)	-0.1001 (0.0975)	-0.1876** (0.0792)
Left-Behind _{jt} × Score _{jt} ^o	0.0571 (0.1268)	0.4467* (0.2423)	-0.1176 (0.1906)	0.3654** (0.1615)	0.1139 (0.1054)	0.6218*** (0.1973)	0.1218 (0.1164)	0.1729* (0.0898)
N	991	575	991	575	1232	334	1232	334
	By Absent Parent							
		Chinese			Math			
		Mother	Father	Both	Mother	Father	Both	
Score _{jt} ^o	0.3489*** (0.0608)	0.3390*** (0.0711)	0.3529*** (0.0607)	0.3836*** (0.0480)	0.3679*** (0.0521)	0.3864*** (0.0477)		
Left-Behind _{jt} (D)	-0.3063*** (0.0996)	-0.2016** (0.0900)	-0.2896*** (0.1014)	-0.0898 (0.0746)	-0.0830 (0.0660)	-0.0797 (0.0758)		
Left-Behind _{jt} × Score _{jt} ^o	0.3667*** (0.1233)	0.2452** (0.1156)	0.3687*** (0.1276)	0.0984 (0.0861)	0.1055 (0.0793)	0.1080 (0.0882)		
N	1566	1566	1566	1566	1566	1566		

Source: RUMiC data, RHS waves 2009, 2010 and 2011. Robust standard errors clustered at the household level in parentheses. All regressions include year fixed effects and the full set of regressors in column V and X of Table 4. MO=Male older sibling; MY=Male younger sibling; FO=Female older sibling; FY=Female younger sibling. "Migrated Before" refers to whether either parent had migrated before the year 2009.

Table 6: Robustness

	Left-Behind Only				Prospective Mig.	
	OLS		Fixed Effects		Chinese	Math
	Chinese	Math	Chinese	Math		
Score _{jt} ⁰	0.3749*** (0.0419)	0.3956*** (0.0362)	0.2831*** (0.0841)	0.3827*** (0.0639)	0.4185*** (0.0620)	0.4082*** (0.0481)
Left-Behind _{jt} (D)	-0.1368** (0.0602)	-0.0457 (0.0538)	-0.2609*** (0.0954)	-0.0799 (0.0726)	-0.1026 (0.0799)	0.0182 (0.0662)
Left-Behind _{jt} × Score _{jt} ⁰	0.1605** (0.0744)	0.0560 (0.0636)	0.3099** (0.1212)	0.0906 (0.0861)	0.1286 (0.0989)	-0.0135 (0.0785)
N	1420	1420	1420	1420	838	838

	Fully Interacted Model			
	OLS		Fixed Effects	
	Chinese	Math	Chinese	Math
Score _{jt} ⁰	0.3852*** (0.0403)	0.3722*** (0.0323)	0.3323*** (0.0688)	0.3671*** (0.0549)
Left-Behind _{jt} (D)	-0.1775* (0.0926)	-0.0227 (0.0855)	-0.3401*** (0.1286)	-0.1259 (0.1362)
Left-Behind _{jt} × Score _{jt} ⁰	0.1605** (0.0722)	0.0921 (0.0614)	0.2858*** (0.1074)	0.1225 (0.0795)
N	1566	1566	1566	1566

	Pr(Left behind _t)		Pr(Left behind _{t+1})		Oldest Sibling	
	Chinese	Math	Chinese	Math	Chinese	Math
Score _{jt} ⁰	0.0242 (0.1373)	-0.0908 (0.1331)	0.3220 (0.3168)	-0.1455 (0.2400)	0.5282*** (0.1277)	0.4979*** (0.0961)
Left-Behind _{jt} (D)					-0.1046 (0.1147)	-0.0887 (0.1050)
Left-Behind _{jt} × Score _{jt} ⁰					0.1252 (0.1417)	0.0894 (0.1266)
Own score	-0.2278 (0.1570)	-0.0922 (0.1591)	-0.0404 (0.3213)	0.4352 (0.3400)		
N	1566	1566	838	838	1189	1189

Source: RUMiC data, RHS waves 2009, 2010 and 2011. Robust standard errors clustered at the household level in parentheses.

All regressions include year fixed effects.

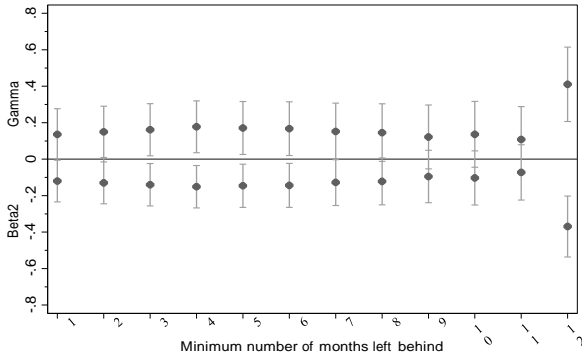
The group "Left-Behind Only" refers to children who are left behind at any year in the panel.

Prospective migrants refer to the estimation of equation 2, in which scores at time t are regressed on an indicator for being Left-Behind at time $t+1$.

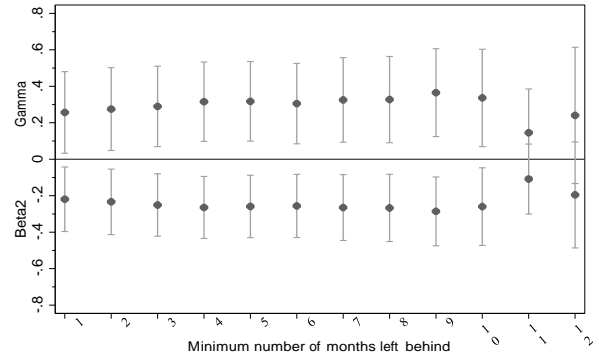
"Oldest Sibling" refers to estimating equation 1 for the sample of oldest siblings.

Appendix

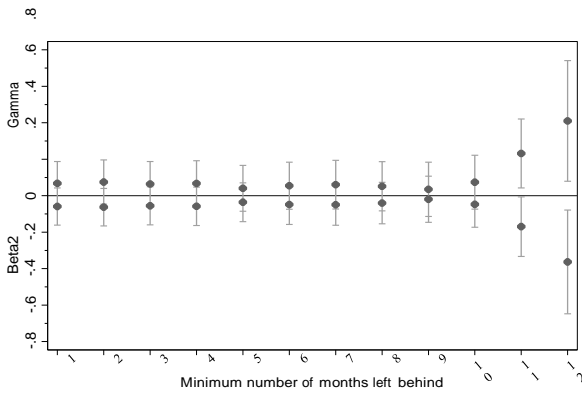
Figure A1: Estimated effects by duration of parents' absence



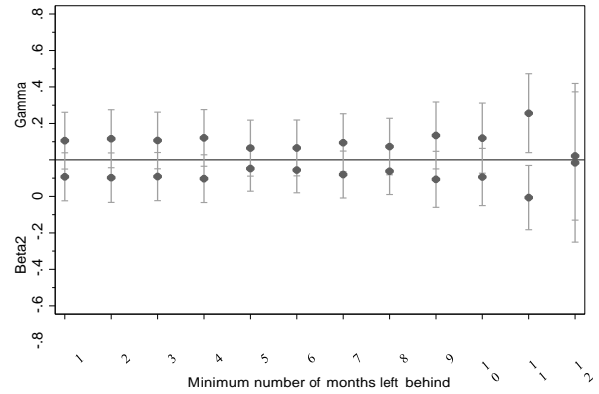
(a) Chinese, OLS



(b) Chinese, Fixed Effects



(c) Math, OLS



(d) Math, Fixed Effects

Table A1: Additional Results

	Balanced Sample				Only 2 Children Households			
	OLS		Fixed Effects		OLS		Fixed Effects	
	Chinese	Math	Chinese	Math	Chinese	Math	Chinese	Math
Score _{jt} ^O	0.4307*** (0.0494)	0.3414*** (0.0665)	0.4550*** (0.0593)	0.3492*** (0.0781)	0.3680*** (0.0496)	0.4047*** (0.0377)	0.2568*** (0.0803)	0.3310*** (0.0479)
Left-Behind _{jt} (D)	-0.2197*** (0.0721)	-0.1990** (0.0801)	-0.1770* (0.0949)	-0.1544 (0.0946)	-0.1911*** (0.0638)	-0.0632 (0.0539)	-0.2856*** (0.1069)	-0.1245* (0.0726)
Left-Behind _{jt} × Score _{jt} ^O	0.2625*** (0.0886)	0.2365** (0.0930)	0.2003* (0.1177)	0.1739 (0.1107)	0.2212*** (0.0789)	0.0802 (0.0636)	0.3390** (0.1366)	0.1605* (0.0851)
N	520	520	520	520	1151	1151	1151	1151
	Without Clustering				Full Interaction Model			
	OLS		Fixed Effects		OLS		Fixed Effects	
	Chinese	Math	Chinese	Math	Chinese	Math	Chinese	Math
Score _{jt} ^O	0.3933*** (0.0379)	0.3782*** (0.0298)	0.3310*** (0.0707)	0.3633*** (0.0549)	0.3852*** (0.0403)	0.3722*** (0.0323)	0.3323*** (0.0688)	0.3671*** (0.0549)
Left-Behind _{jt} (D)	-0.1203** (0.0554)	-0.0586 (0.0462)	-0.2194** (0.0879)	-0.0926 (0.0684)	-0.1775* (0.0926)	-0.0227 (0.0855)	-0.3401*** (0.1286)	-0.1259 (0.1362)
Left-Behind _{jt} × Score _{jt} ^O	0.1403** (0.0686)	0.0720 (0.0549)	0.2564** (0.1110)	0.1057 (0.0808)	0.1605** (0.0722)	0.0921 (0.0614)	0.2858*** (0.1074)	0.1225 (0.0795)
N	1566	1566	1566	1566	1566	1566	1566	1566
	Initial Performance Above Median				Initial Performance Below Median			
	OLS		Fixed Effects		OLS		Fixed Effects	
	Chinese	Math	Chinese	Math	Chinese	Math	Chinese	Math
Score _{jt} ^O	0.1712*** (0.0412)	0.2110*** (0.0355)	0.1241* (0.0708)	0.2187*** (0.0745)	0.3260*** (0.0381)	0.2690*** (0.0412)	0.4121*** (0.0645)	0.3663*** (0.0707)
Left-Behind _{jt} (D)	-0.0488 (0.0515)	0.0243 (0.0462)	-0.1952** (0.0833)	-0.1637** (0.0804)	-0.1670** (0.0739)	-0.1151* (0.0587)	-0.1970** (0.0907)	-0.0720 (0.0893)
Left-Behind _{jt} × Score _{jt} ^O	0.0595 (0.0613)	-0.0258 (0.0528)	0.2303** (0.0970)	0.1913** (0.0948)	0.2029** (0.0983)	0.1523** (0.0746)	0.2377* (0.1222)	0.0698 (0.1089)
N	770	799	770	799	796	767	796	767

Source: RUMiC data, RHS waves 2009, 2010 and 2011. Robust standard errors clustered at the household level in parentheses. All regressions include year fixed effects and the full set of regressors in columns V and X of Table 4.

Table A2: Performance in Chinese and Math, OLS Results, varying set of observable characteristics

	Chinese				Math			
	I	II	III	IV	V	VI	VII	VIII
Left-Behind $_{jt}$ × Score $^O_{jt}$	0.132* (0.071)	0.132* (0.071)	0.146** (0.072)	0.140* (0.072)	0.053 (0.060)	0.052 (0.060)	0.074 (0.061)	0.072 (0.061)
Score $^O_{jt}$	0.406*** (0.040)	0.408*** (0.039)	0.401*** (0.039)	0.393*** (0.039)	0.392*** (0.032)	0.391*** (0.032)	0.382*** (0.031)	0.378*** (0.032)
Left-Behind $_{jt}$ (D)	-0.113* (0.058)	-0.119** (0.058)	-0.129** (0.058)	-0.120** (0.059)	-0.045 (0.051)	-0.050 (0.051)	-0.066 (0.052)	-0.059 (0.052)
Male (D)	-0.006 (0.005)	-0.006 (0.005)	-0.007 (0.005)	-0.008 (0.005)	0.000 (0.006)	0.000 (0.006)	-0.002 (0.006)	-0.001 (0.006)
Age	-0.003* (0.002)	-0.003* (0.002)	-0.003 (0.002)	-0.004** (0.002)	-0.002 (0.002)	-0.002 (0.002)	-0.002 (0.002)	-0.003* (0.002)
Age at entry	-0.008* (0.004)	-0.008* (0.004)	-0.007* (0.004)	-0.010** (0.004)	-0.009** (0.004)	-0.009** (0.004)	-0.008** (0.004)	-0.011*** (0.004)
Boarding school	-0.007 (0.006)	-0.007 (0.006)	-0.006 (0.006)	-0.006 (0.006)	-0.008 (0.007)	-0.007 (0.007)	-0.006 (0.007)	-0.004 (0.007)
High quality school	0.034*** (0.006)	0.033*** (0.006)	0.031*** (0.006)	0.028*** (0.006)	0.037*** (0.007)	0.036*** (0.007)	0.033*** (0.007)	0.031*** (0.006)
Grade	-0.002 (0.002)	-0.002 (0.002)	-0.002 (0.002)	-0.001 (0.002)	-0.003 (0.002)	-0.003 (0.002)	-0.003 (0.002)	-0.002 (0.002)
Father is employee		0.005 (0.006)	0.003 (0.006)	0.002 (0.006)		0.006 (0.008)	0.003 (0.007)	0.002 (0.007)
Father is self-employed		-0.005 (0.009)	-0.006 (0.008)	-0.004 (0.008)		0.001 (0.009)	-0.000 (0.009)	0.000 (0.009)
Mother is employee		0.015** (0.007)	0.016** (0.007)	0.012* (0.007)		0.011 (0.007)	0.012* (0.007)	0.009 (0.007)
Mother is self-employed		0.011 (0.012)	0.011 (0.012)	0.009 (0.012)		0.016 (0.014)	0.017 (0.014)	0.012 (0.014)
Household size			-0.004* (0.002)	-0.003 (0.003)			-0.007*** (0.002)	-0.005** (0.003)
Household financial asset			-0.001 (0.001)	-0.001 (0.001)			-0.001 (0.001)	-0.001 (0.001)
Household land size			-0.037*** (0.011)	-0.028** (0.012)			-0.034*** (0.011)	-0.030** (0.012)
Lag village population				0.000 (0.002)				0.001 (0.002)
Lag village expenditure on education				0.005 (0.021)				-0.018 (0.020)
Lag village number of teachers				0.015*** (0.004)				0.013** (0.006)
Lag village number of students				-0.002** (0.001)				-0.001 (0.001)
R^2	0.32	0.32	0.33	0.35	0.29	0.29	0.30	0.32
N	1566	1566	1566	1566	1566	1566	1566	1566

Source: RUMiC data, RHS waves 2009, 2010 and 2011. Robust standard errors clustered at the household level in parentheses. Column IV and VIII include province fixed effects. All regressions include year fixed effects.

Table A3: Performance in Chinese and Math, Fixed Effect Results, varying set of observable characteristics

	Chinese				Math			
	I	II	III	IV	V	VI	VII	VIII
Left-Behind $_{jt} \times$ Score $_{jt}^O$	0.249** (0.115)	0.253** (0.113)	0.259** (0.113)	0.256** (0.114)	0.103 (0.081)	0.108 (0.080)	0.111 (0.080)	0.106 (0.079)
Score $_{jt}^O$	0.331*** (0.075)	0.331*** (0.072)	0.330*** (0.072)	0.331*** (0.072)	0.364*** (0.053)	0.359*** (0.052)	0.360*** (0.052)	0.363*** (0.053)
Left-Behind $_{jt}$ (D)	-0.208** (0.090)	-0.216** (0.089)	-0.221** (0.089)	-0.219** (0.090)	-0.086 (0.068)	-0.094 (0.068)	-0.097 (0.068)	-0.093 (0.067)
Boarding school	0.008 (0.009)	0.007 (0.009)	0.007 (0.009)	0.006 (0.009)	0.004 (0.011)	0.003 (0.011)	0.003 (0.011)	0.002 (0.011)
High quality school	0.023** (0.011)	0.023** (0.011)	0.022** (0.011)	0.021** (0.011)	0.020 (0.012)	0.020 (0.012)	0.021* (0.012)	0.021* (0.012)
Grade	-0.002 (0.002)	-0.002 (0.002)	-0.002 (0.002)	-0.002 (0.002)	-0.005 (0.003)	-0.005 (0.003)	-0.005 (0.003)	-0.005 (0.003)
Father is employee		0.012 (0.014)	0.012 (0.014)	0.013 (0.014)		0.012 (0.013)	0.015 (0.013)	0.016 (0.014)
Father is self-employed		-0.003 (0.019)	-0.003 (0.019)	-0.002 (0.019)		0.015 (0.020)	0.020 (0.020)	0.021 (0.020)
Mother is employee		0.008 (0.015)	0.009 (0.015)	0.009 (0.015)		0.002 (0.013)	-0.002 (0.014)	-0.002 (0.014)
Mother is self-employed		-0.023 (0.026)	-0.021 (0.025)	-0.022 (0.025)		-0.021 (0.024)	-0.024 (0.025)	-0.027 (0.024)
Household size			0.004 (0.010)	0.003 (0.011)			-0.007 (0.014)	-0.007 (0.013)
Household financial asset			-0.002 (0.002)	-0.002 (0.002)			-0.000 (0.002)	-0.000 (0.002)
Household land size			-0.017 (0.014)	-0.015 (0.014)			-0.036** (0.016)	-0.038** (0.016)
Lag village population				0.003 (0.007)				-0.008 (0.007)
Lag village expenditure on education				0.078 (0.088)				0.026 (0.057)
Lag village number of teachers				0.021 (0.017)				0.021 (0.017)
Lag village number of students				-0.002 (0.001)				0.001 (0.001)
R^2	0.27	0.28	0.28	0.29	0.22	0.22	0.23	0.23
N	1566	1566	1566	1566	1566	1566	1566	1566

Source: RUMiC data, RHS waves 2009, 2010 and 2011. Robust standard errors clustered at the household level in parentheses.
 Column IV and VIII include province fixed effects.
 All regressions include year fixed effects.

References

- Abbey, C. and R. Dallos (2004). The experience of the impact of divorce on sibling relationships: A qualitative study. *Clinical Child Psychology and Psychiatry* 9(2), 241–259.
- Akgüc, M., C. Giulietti, and K. F. Zimmermann (2014). The RUMiC longitudinal survey: Fostering research on labor markets in China. *IZA Journal of Labor and Development* 3(5).
- Altonji, J. G., S. Cattan, and I. Ware (2010). Identifying sibling influence on teenage substance use. *NBER Discussion Paper 16508*.
- Angrist, J. D., S. M. Dynarski, T. J. Kane, P. A. Pathak, and C. R. Walters (2012). Who benefits from kipp? *Journal of Policy Analysis and Management* 31(4), 837–860.
- Antman, F. M. The impact of migration on family left behind. In A. F. Constant and K. F. Zimmermann (Eds.), *International Handbook on the Economics of Migration*, pp. 293–308. Cheltenham: Edward Edgar.
- Antman, F. M. (2011a). The intergenerational effects of paternal migration on schooling and work: What can we learn from children's time allocations? *Journal of Development Economics* 96(2), 200–208.
- Antman, F. M. (2011b). International migration and gender discrimination among children left behind. *The American Economic Review* 101(3), 645–649.
- Antman, F. M. (2012). Gender, educational attainment, and the impact of parental migration on children left behind. *Journal of Population Economics* 25(4), 1187–1214.
- Arnold, J., A. Levine, and G. Patterson (1975). Changes in sibling behavior following family intervention. *Journal of Consulting and Clinical Psychology* 43(5), 683–688.
- Behrman, J. R., R. A. Pollak, and P. Taubman (1982). Parental preferences and provision for progeny. *The Journal of Political Economy* 90(1), 52–73.
- Bertrand, M. and S. Mullainathan (1999). Is there discretion in wage setting? A test using takeover legislation. *The Rand Journal of Economics* 30(3), 535–554.
- Björklund, A. and K. G. Salvanes (2010). Education and family background: Mechanisms and policies. In E. A. Hanushek, S. Machin, and L. Woessmann (Eds.), *Handbook of the Economics of Education*, Volume 3. Amsterdam: North Holland.
- Black, S. E. and P. J. Devereux (2011). Recent developments in intergenerational mobility. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 4B. Amsterdam: Elsevier.
- Chen, X., Q. Huang, S. Rozelle, Y. Shi, and L. Zhang (2009). Effect of migration on children's educational performance in rural China. *Comparative Economic Studies* 51(3), 323–343.

- Edmonds, E. V. (2006). Understanding sibling differences in child labor. *Journal of Population Economics* 19(4), 795–821.
- Hanson, G. H. and C. Woodruff (2003). Emigration and educational attainment in Mexico. Unpublished manuscript, University of California at San Diego.
- Hanushek, E. A. (1992). The trade-off between child quantity and quality. *Journal of Political Economy* 100(1), 84–117.
- Haynie, D. L. and S. McHugh (2003). Sibling deviance: In the shadows of mutual and unique friendship effects? *Criminology* 41(2), 355–392.
- Ilahi, N. (2001). *Children's Work and Schooling: Does Gender Matter?: Evidence from the Peru LSMS Panel Data*. Number 2745. World Bank, Latin American and the Caribbean Region, Gender Sector Unit.
- Kaestner, R. (1997). Are brothers really better? Sibling sex composition and educational achievement revisited. *Journal of Human Resources* 32(2), 250–84.
- Kaestner, R. and O. Malamud (2014). Self-selection and international migration: New evidence from Mexico. *Review of Economics and Statistics*, 96(1), 78–91.
- Kong, T. and X. Meng (2010). The educational and health outcomes of the children of migrants. In X. Meng and C. Manning (Eds.), *The Great Migration: Rural-Urban Migration in China and Indonesia*, pp. 93–116. Cheltenham: Edward Edgar.
- Levison, D. and K. S. Moe (1998). Household work as a deterrent to schooling: An analysis of adolescent girls in Peru. *The Journal of Developing Areas*, 339–356.
- Liang, Z. and Y. P. Chen (2007). The educational consequences of migration for children in China. *Social Science Research* 36(1), 28–47.
- Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies* 60(3), 531–542.
- Metzler, J. and L. Woessmann (2012). The impact of teacher subject knowledge on student achievement: Evidence from within-teacher within-student variation. *Journal of Development Economics* 99(2), 486–496.
- Meyerhoefer, C. D. and C. Chen (2011). The effect of parental labor migration on children's educational progress in rural China. *Review of Economics of the Household* 9(3), 379–396.
- Moraga, J. F.-H. (2011). New evidence on emigrant selection. *The Review of Economics and Statistics* 93(1), 72–96.
- NBS China (2013). Accessed at http://www.stats.gov.cn/was40/gjtjj_en_detail.jsp?searchword=migrants&channelid=9528&record=3 on 12 December 2013.
- Nicoletti, C. and B. Rabe (2014). Sibling spillover effects in school test scores. *Unpublished*

Manuscript.

- Oettinger, G. S. (2000). Sibling similarity in high school graduation outcomes: Causal interdependency or unobserved heterogeneity? *Southern Economic Journal* 66(3), 631–648.
- Qureshi, J. A. (2011). Additional returns to investing in girls' education: Impact on younger sibling human capital.
- Rees, D. I. and J. J. Sabia (2009). The effect of breast feeding on educational attainment: Evidence from sibling data. *Journal of Human Capital* 3(1), 43–72.
- Schnitzlein, D. D. (2014). How important is the family? Evidence from sibling correlations in permanent earnings in the USA, Germany, and Denmark. *Journal of Population Economics* 27(1), 69–89.
- Solon, G. (1999). Intergenerational mobility in the labor market. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 3. Amsterdam: Elsevier.
- Wang, S. X. (2012). The effect of parental migration on the educational attainment of their left-behind children in rural China. Unpublished manuscript, Central University of Finance and Economics Beijing.
- Wei, S.-J. and X. Zhang (2011). The competitive saving motive: Evidence from rising sex ratios and savings rates in China. *Journal of Political Economy* 119(3), 511–564.
- Zimmerman, D. J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *Review of Economics and Statistics* 85(1), 9–23.