The Effect of School Transfers on Academic and Non-academic Performance of Rural-to-Urban Migrant Children in China

Xuche Gong, PhD student, Department of Agricultural, Food, and Resource Economics, Michigan State University, East Lansing, MI.
Email:xuche_gong@qq.com

Yan Yuan, Associate Professor, Southwestern University of Finance and Economics, Chengdu, China 610074. Email: yanyuan.tamu@gmail.com

Preliminary draft

Please do not cite this version of the paper without prior permission from the authors

Selected Paper prepared for presentation at the 2017 Agricultural & Applied Economics Association Annual Meeting, Chicago, Illinois, July 30-August 1

The Effect of School Transfers on Academic and Non-academic Performance of Rural-to-Urban Migrant Children in China

Abstract: China has seen an increasing trend that rural children migrate to cities with their parents in recent years. They largely lagged behind their urban counterparts in their educational and health outcomes due to their disadvantaged socioeconomic status and *Hukou* constraint. Using propensity score matching method and data from 2013-2014 China Education Panel Survey, we investigate whether higher frequency of school transfers contribute to their sluggish developments. We find that although school transfers in primary school years did not harm migrant children's cognitive skills, they have a strong impact on the probability of grade retention. In extreme cases when children experienced 3 or more school transfers, it significantly reduced migrant children's willingness to attend college in the future, and increased their depression level.

Keywords: School transfer; Migrant children; Academic performance; China

I. Introduction

Rural-to-urban migration is a long-standing phenomenon in modern China where migrant workers travel back and forth between destination and their origins, leaving their children and other family members in countryside. However, it has seen a recent change in the trend that more children migrate together with their parent(s). According to the Report of Migrant Children in China (2014), about 29 million Chinese rural-registered children were living in cities with their migrating parents, making up a fifth of urban children. However, compared with their local urban counterparts, migrant children are faced with many obstacles in pursing personal development in cities.

The *hukou* system and insufficient public education resources are the most discussed factors in literature that constrain the development of migrant children. Although all school-age children in China are entitled to a free and compulsory 9-year education by law, funding for elementary education received by the local government is determined by the number of children with local *hukou,* and it is not portable across counties. Therefore, local government lack the incentives and resources to accommodate educational needs of migrant children, who usually do not have urban *hukou*. As a result, a significant part of migrant children is excluded from urban public education system and can only attend the so-called "migrant schools". These migrant schools serve exclusively the educational needs of migrant children in cities and are much inferior in education quality and school environment to public schools (Chen and Feng, 2013). Studies have documented, compared with enrolling in public schools, enrolling in migrant schools has an adverse effect on migrant children's social adaptiveness and academic performance (Chen and Feng, 2013; Lai *et al*., 2014; Zeng and Li, 2007).

Besides, socioeconomically disadvantaged status of migrant children also hinders their development. In a survey conducted by Xie *et al*. (2011), about 51.3% of 2261 migrant workers in the Pearl River Delta (PRD) list school fees as the most important factor determining the school choice of their children, while 32% chooses school quality as the most important factor. Therefore, migrant families and their children are more likely to be faced with budget constraints to provide necessary resources for educational and health purposes. In this regard, family factors may be important in determining migrant children's health and academic outcomes, which received little attention in extant literature. In this paper, we investigate whether household characteristics have any impacts on their cognitive skills and health outcomes. We pay special attention to one characteristics of migrant children, i.e., the high frequencies of school transfers.

Migrant workers in China usually take jobs with high mobility, and consequently, migrant children often changed schools as their parents changed jobs and residences. In the survey of Xie *et al*. (2011), only about 18.1% of migrant workers in PRD have never changed a job within the previous year, while 44.3% of them have changed jobs at least 3 times. This is not unique to Pearl River Delta. As shown in Figure 1, according to the 2013-2014 China Education Panel Survey (CEPS), 46.1% of rural-to-urban migrant children changed schools at least once in their primary school years, and 10.9% of them have changed schools at least three times. As a comparison, only 22% of local urban children have ever changed a school and only 3.1% of them have changed schools at least 3 times during their primary school years. A strand of literature found that frequent school transfers reduce children’s cognitive skills and academic performance (Temple and Reynolds, 1999; Mehana and Reynolds, 2003; Burkam *et al*., 2009) and increase the probability of grade retention

(Ginsburg *et al*., 2011) and that of dropping out of school (Rumberger and Larson, 1998; South *et al*., 2007). If school transfer also has an adverse effect on migrant children's development, the potential benefit of investing huge resources to improve the school quality of migrant student may be discounted, as it may not change the education motivation of migrant children's parents and thus have no impact on children's school transfer behaviors.

Using the 2013-2014 China Education Panel Survey, we study the effect of school transfers on migrant children's academic performance and health outcomes. We depart from the binary treatment case since school transfers occur at different frequencies, which can be considered as different levels of treatment. Thus we not only compare the outcomes of migrant children who ever transferred schools to those who have never transferred, but also compare the outcomes of migrant children who transferred schools more to those transferred less. We use propensity score matching (PSM) method to address the potential selection bias and reverse causality problem. Since we observe the major variables that simultaneously affect selection and outcomes, we assume that child outcomes and selection of school transfers are independent conditional on these observables (Conditional Independence Assumption – CIA). We examine the appropriateness of the CIA, and further conduct a sensitive test about how the inference of our estimated treatment effects would change if different scales of hidden bias exist.

We found that although school transfers in primary school years did not harm migrant children's cognitive skills, experiencing 3 or more school transfers significantly reduced their willingness to attend college in the future, and significantly increased their depression level. Experiencing school transfer ever, however, increases their probability to repeat a grade, and the probability increased with number of school transfers. We also find that transferring school just once has a statistically significantly positive impact on migrant children's cognitive skills and their willingness to attend college.

The paper unfolds as follows: In section 2 we review the main findings of the school transfer literature. Section 3 introduces our model specification. Section 4 presents our data and variables. Sections 5 exhibits our estimation results and section 6 concludes.

II. Literature review

School mobility are believed to be a risk factor because it introduces discontinuity in learning environment (Bronfenbrenner, 1979). Astone and Mclanahan (1994) put forward three mechanisms that residential mobility might harm children development, which also apply for the case of school mobility: First, children who often change schools may miss educational material, thereby lowering their school performance; Moreover, children (and parents) who are new to a new community have less information about the school system and thus are less able to take full advantage of the resources in a particular school than children who have lived in the community for a long time; In addition, residential mobility may undermine children's relationships with teachers and peers, and children attending a new school may feel socially isolated or marginalized and seek disengaged from the education process. Using data from the High School and Beyond Study (HSB), they find that as much as 30% of the difference in the risk of dropping out between children from stepfamilies and children from intact families can be explained by differences in residential mobility, which usually result in school transfer.

Early studies that directly study the impact of school transfers on child development also find a negative relationship. Using the National Educational

Longitudinal Survey data, Rumberger and Larson (1998) find that after controlling for other predictors, students who made even one nonpromotional school change between the eighth and twelfth grades were twice as likely to not complete high school as students who did not change schools. Using the Chicago Longitudinal Study, Temple and Reynolds (1999) find that school changes could explain about half of the lagged performance of children who changed schools frequently between kindergarten and seventh grade. In a meta-analysis, Mehana and Reynolds (2003) evaluated the effects of school mobility on student achievement in the elementary grades from 26 studies dated between 1975 and 1994. They find that the average achievement level of mobile students exceeded that of about 40% of the non-mobile students, which is equivalent to a 3-4 month performance disadvantage in achievement. Some later studies also find that school transfers are negatively correlated with academic performance (Gruman *et al.*, 2008; Xu *et al.*, 2011), classroom participation (Gruman *et al.*, 2008), the possibility of dropping out of school (South *et al.*, 2007), cognitive skills (Burkam *et al.*, 2009), and the probability of grade retention (Ginsburg *et al.*, 2011), and the effects are larger and more significant for socioeconomically disadvantaged children (South *et al.*, 2007; Burkam *et al.*, 2009; Xu *et al.*, 2011).

However, some studies show that the strength and pattern of associations between educational outcomes and school mobility are likely to be influenced by a series of individual, family, household and school factors. Alexander *et al.* (2001) find that after 5 years in school, children who changed school had lower test scores and marks, had an elevated risk of grade retention, and were more likely to receive special education service. But most of those differences fell short of significance when controls were introduced for first-grade measures of school performance and for background characteristics. Preibesh and Downey (1999) and Strand and Demie

(2006) also find that most negative effects of school transfers on children's education performance are due to preexisting differences between two groups. Ginsburg *et al*. (2011) further point out that the effect of school transfers may be context based. They find no relationship between changes in school and competency in numeracy and literacy in South Africa.

The inconsistence of literature highlights the importance of excluding the impacts that simultaneously influence school transfers and child development. Propensity score matching method is a possible way to achievement this goal as we will state below.

III. Empirical Methodology

In this section, we introduce the methodology we use to perform our empirical analysis. We apply Propensity score matching (PSM) to estimate the effects of school transfers on health and academic outcomes of migrant children. PSM has become a very popular approach to estimate causal treatment effect. It applies for all situations where one has a treatment, a group of treated individuals and a group of untreated individuals (Caliendo and Kopeing, 2008).

Formally, in the binary treatment case, we assume that there is a variable T_i indicating treatment, which equals to one if individual i belongs to the treatment group and zero otherwise. If we define the child outcomes as Y_{0i} and Y_{1i} for the associated states 0 and 1, then the treatment effect for individual i can be written as:

$$t_i = Y_{0i} - Y_{1i}. \tag{1}$$

Since economists care about the policy effects on the intended group, the parameter receiving most attention in the literature is the average treatment effect on the treated:

$$t_{ATT} = E(t|T = 1) = E[Y_1|T = 1] - E[Y_0|T = 1]. \tag{2}$$

However, we do not know t_i for everyone since we can only observe either Y_{0i} or Y_{1i}, thus we cannot observe $E[Y_0|T=1]$. Substituting $E[Y_0|T=1]$ with $E[Y_0|T=0]$ will lead to self-selection bias since factors that determine the treatment decision is also likely to determine the outcomes, and thus the outcomes of individuals from the treated and untreated groups would usually differ even in the absence of treatment. PSM provides us an alternative way to find a comparable untreated group for the treated group.

The basic idea of PSM is to find a group of untreated individuals that are similar to the treated ones in all relevant pretreatment characteristics. The propensity score is defined as the conditional probability of receiving the treatment given pretreatment characteristics X:

$$p(\mathrm{X}) \equiv \Pr(T=1|X) = E(T|X). \quad (3)$$

Based on the following assumptions, we can get a consistent PSM estimator (see also (Caliendo and Kopeing, 2008; Becker and Ichino, 2002; Imbens, 2000):

$$t_{ATT}^{PSM} = E_{p(\mathrm{X})|T=1}\{E[Y_1|T=1, p(\mathrm{X})] - E[Y_0|T=0, p(\mathrm{X})]\}. \quad (4)$$

Assumption 1. Unconfoundeness or conditional independence assumption:

$$Y_1, Y_0 \perp T|X \quad (5)$$

Assumption 2. Common support or overlap condition:

$$0 < P(T=1|X) < 1. \quad (6)$$

Assumption 1 implies that selection is solely based on observables and all variables that influence assignment and potential outcomes simultaneously are observed by the researcher (Caliendo and Kopeing, 2008). Using the exact set of the observed variables as required for CIA to hold is a necessary step for the unbiased estimation of treatment effects. This is a strong assumption that has to be justified by the data quality at hand. In the later parts we will examine the appropriateness of the

CIA in our study.

Assumption 2 ensures that individuals with the same X values have a positive probability of being both participants and nonparticipants (Heckman *et al*., 1999). The method of matching assumes that, given X, some unspecified randomization device allocates people to treatment (Heckman and Navarro-Lozano, 2004).

The method of matching with a known conditioning set does not require separability of outcome or choice equations, exclusion restrictions, or the adoption of specific functional forms of outcome equations that are common in conventional selection methods and conventional instrumental variable formulations (Heckman and Navarro-Lozano, 2004). It does not require exogeneity of conditional variables, either. Lechner (2008) showed that it does not matter when some of control variables may be influenced by the treatment as long as the usual formulation of the CIA holds.

IV. Data and variables

The data used in this paper comes from the 2013-2014 China Education Panel Survey (CEPS), the latest available dataset. CEPS is designed to investigate the linkage between individuals' educational outcomes and multiple contexts of families, school processes, communities and social structure, and further studies the effects of educational outcomes during people's life course. It starts with two cohorts – the 7th and 9th graders in the 2013-2014 academic year, randomly selecting a school-based, nationally representative sample of approximately 20000 students in 438 classrooms of 112 schools in 28 county-level units in mainland China. It contains detailed information on children, their home and school environments, and demographic information. With a special care on migrant children, CEPS includes an overweight sample of migrant children from 13 counties/districts with a large part of migrant population and detailed recorded their migration history, which makes it the best dataset of our study goal.

Rural-to-urban migrant children are defined as children with a rural *hukou* and currently live in cities other than the county at which their *hukou* registered. Although CEPS has two student cohorts – the 7th graders and 9th graders, we only include the 7th graders in our analysis sample since 9th graders who are not allowed to attend the urban local high school may have already returned to their rural hometown and are not in the dataset, which may lead to a severe sample selection problem.

School transfer is measured at numbers. We depart from the binary treatment case since school transfers occur at different frequencies, which can be considered as different levels of treatment. We create four categories of school transfers (never, once, twice, more than twice). Since the multinomial treatment model is computationally burdensome, we follow the advice of Lechner (2002) to estimate a series of binomial models instead of a multinomial treatment model. Therefore, besides the binomial model of >=1 vs. 0, we also estimate 6 other binomial models: 1 vs. 0, 2 vs. 0, >=3 vs. 0, 2 vs. 1, >=3 vs1, >=3 vs. 2.

We choose children's cognitive skills to measure the academic outcomes of migrant children. Past literature has well established the importance of cognitive and non-cognitive skills accumulated in childhood on determining an array of outcomes, such as schooling, occupation, and income (Cameron and Heckman, 1998; Cameron and Heckman, 2001; Heckman *et al*., 2006). In CEPS all 7^{th} graders are require to answer 20 questions regarding their language skills, graph analysis ability, and computation and logistic ability. The number of questions that a child answers correctly was recorded as his/her final score. CEPS further standardized the test scores of all 7^{th} graders to have a distribution of mean 0 and standard deviation 0.873.

We choose whether the migrant child has experienced any grade retention in his/her primary school years to measure whether the child has failed to follow the grade curriculum. Also, a strand of literature show that grade retention in early school years is highly correlated with later school dropout and low education attainment (Jacob and Lefgren, 2009; Manacorda, 2012; Tafreschi and Thiemann, 2016), thus it may also be a measure of the risk of school dropout.

We choose children's willingness to attend college and their depression level to measure children's education motivation and emotional status. In CEPS, all students were asked to answer 5 questions regarding their mental status last week[1]. Answers to each question can take 5 values: 1, Almost every day; 2, 2-3 times a week; 3, 2-3 times a month; 4, once a month; 5, never. The principal factor analysis on those questions gets only one factor whose eigenvalue is larger than 1, and a common interpretation of the factor is the depression level of the respondent. A higher score of depression level means the child was more depressed.

We divide variables that simultaneously affect school transfers and child outcomes into three categories: child variables, family income variables, and family education variables. Child variables including the age of the child and the age when he/she started his/her primary school, gender, number of siblings, ethnics, whether the child has attended a kindergarten before primary school years, whether the child has ever in a bad disease before primary school years and whether the child migrated across province. Family income variables include parents' occupations, parents-report family income level before primary school and whether the family is currently receiving the minimum subsistence allowance. Family education variables include parents' education level, parents' relationship with each other, whether the father often

[1] The 5 questions are: feel upset in last 7 days; feel depressed in last 7 days; feel unhappy in last 7 days; feel life is meaningless in last 7 days; feel sad in last 7 days.

gets drunk and the strictness level of parents on the child[2]. Since we control a very exhausted set of variables that both have an effect on school transfers and outcome variables, we expect to have little unobservable heterogeneity left and that the CIA assumption holds. Furthermore, in later parts, we conduct sensitive analysis to test how sensitive our estimates are to potential failure of CIA. After dropping observations with missing values on these variables, we get a final sample of 989 children. Description of variables is listed in Table 1.

V. Estimation and results

1. Matching procedure

The first two concerns when estimating the propensity score are the model used for estimation and variables to be included in the model. According to Lechner (2002), a binary probit model and a multinomial probit model yield similar results. Following it, we then proceed by estimating a binary probit model.

The choice of the variables builds on the CIA, requiring that the outcome variables should be independent of treatments conditional on the propensity score (Caliendo and Kopeinig, 2008). Only variables simultaneously influence the participation decision and outcomes, and only variables not affected by the participation (and its participant) should be included. To ensure this we only choose variables that are likely to be fixed over time and variables that are measured before primary school years to estimate the propensity score. Table 1 present the descriptive statistics of variables included.

Table 2 presents the results of the probit model for the propensity score estimations. Children who are older, having more siblings, and who migrate across

[2] The strictness level of parents on the child is measure by 8 questions: Are you strict on this child about his/her: Homework and examination; Behavior at school; Attendances at school everyday; Time when he/she get home everyday; Whom he/she make friends with; His/her dress style; Time he/she spends on the Internet; Time he/she spends on watching TV. The principal factor analysis on those questions gets only one factor whose eigenvalue is larger than 1, we interpret it as parental strictness on this child.

provinces were more likely to experience more school transfers, while children who are *Hans* were less likely to experience school transfers. Children whose mother are taking jobs with high skills experienced fewer school transfers, while children from families currently receiving the minimum subsistence allowance were more likely to experience school transfers. Among family education motivation variables, children whose mother are less educated were less likely to experience school transfers than those whose mother are college educated, children who have strict parents were less likely to experience school transfers than children who have less strict parents.

The next step in calculation of the propensity score estimator is the choice of a matching algorithm. Asymptotically, all matching algorithms should yield the same results. However, in small samples the choice of matching algorithms can be important (Heckman *et al*., 1997a), where usually a trade-off between bias and variance arises. Caliendo and Kopeinig (2008) suggest trying a number of approaches. Hence, we implement six matching algorithms (i.e. one-to-one nearest neighbor, kernel matching, local linear, spline matching and radius matching with caliper levels 0.1 and 0.01).

Testing the statistical significance of treatment effects and computing their standard errors is not straightforward since the estimation steps precede the matching process add variations. We used bootstrapping to address this problem, which we repeated 800 times for each matching algorithms to derive the bootstrapped standard errors of ATT. As Abadie and Imbens (2006) show that bootstrap variance estimator is invalid for nearest neighbor matching, we did not calculate the bootstrap estimator for this algorithm.

Table 3 presents the estimated ATT's for each model and for each outcome variable. We find that for each model, all matching algorithms yield similar results in terms of magnitude and statistical significance. The results of the unmatched sample are also consistent with the results of the matched sample in most cases.

For cognitive skills, we find no evidence that migrant children who have ever transferred schools have significantly lower cognitive skills than migrant children who have never transferred schools. Also, there is no evidence that among children who have ever transferred school, those who transferred more have significantly lower cognitive skills than those who transferred less. What is more interesting is that we find migrant children who have transferred school just once have significantly higher cognitive skills than migrant children who have never transferred a school.

For grade retention, there is strong evidence that migrant children who have ever transferred schools in primary school years have a significantly higher probability of experiencing grade retention than migrant children who have never transferred a school. The probability goes up with transfer times.

For education expectation, we find that migrant children who have transferred schools 3 or more times have significantly lower willingness to attend college than children who have never transferred a school and children who have transferred schools no more than 2 times. Also, similar to the case of cognitive skills, migrant children who have transferred school just once have a significantly higher willingness to attend college than children who have never transferred a school.

We find that migrant children who have transferred schools 3 or more times are significantly more depressed than migrant children who have never transferred a school and who have transferred school just once. Also, migrant children who have ever transferred school twice have significantly higher depression level than children who have just transferred school once.

The results above reveal that although school transfers do not have a detrimental effect on migrant children's cognitive skills, they do increase the probability of experiencing grade retention and transferring school 2 or more times have some negative effects on children's willingness to attend college and their depression level. However, compared with never transferring a school, transferred school just once have a significantly positive effect on migrant children's cognitive skills and significantly increased their willingness to attend college. To further test the credibility of these results, we conducted a sensitive test of our statistical inference in later sections of the paper.

2. Common support

It is important to check the overlap and the region of common support for the treated and untreated group. First, a visual analysis of the density distributions of the propensity scores is shown in Figure 2. The bottom half of each graph shows the propensity score distribution for the untreated, while the upper half refers to the treated individuals. Problems would arise if the distributions did not overlap. We imposed the common support using the minima and maxima comparison. The basic criterion of this approach is to delete all observations whose propensity score is smaller than the minimum and larger than the maximum in the opposite group. Hence, we removed from our analysis the treated individuals who fall outside the common support region. Table 4 contains the number of observations lost in each model and the propensity score regions after the common support imposition. The number of lost observations in most cases is quite low. Specifically, we lost only a very small fraction (0.2%) of the sample in a vast majority of the models.

3. Matching quality

In this section, we further check whether the matching process is able to balance the distribution of our conditional variables. We use the standardized bias (SB) measure proposed by Rubin (1991) to check if there are differences remaining after conditioning on the propensity score. For each covariate *X*, the SB is the difference of the sample means in the treated and matched comparison sub-samples as a percentage of the square root of the average of the sample variances in both groups (Caliendo and Kopeinig, 2008). For abbreviation, we calculated the means of the SB (MSB) before and after matching by model and matching algorithm. The first row of the unmatched section in Table 5 shows that before matching, the overall bias lies between 9.08% and 17.03%. The bias is largely reduced after matching, especially for the kernel estimator and the radius estimator (caliper = 0.01). These results clearly indicate that the matching procedure is able to balance the characteristics between the treated and matched untreated groups.

Another method is to calculate the pseudo-R2 to test if there are systematic differences in the distribution of covariates after matching. The pseudo-R2 indicates how well the regressors explain the participation probability. If the distribution of covariates of the treated and untreated groups in the matched sample is well balanced, the pseudo-R2 should be low. As shown in Table 5, this is true for our matching estimators.

Finally, we perform a likelihood ratio test on the joint significance of all regressors. Before matching, the test should be accepted. A rejection of the test after matching reflects a good balancing of the covariates. As exhibited in Table 5, this is also true in all of our cases.

4. Sensitive tests for hidden bias

Propensity score matching estimators are based on the assumption that conditional on propensity score, selecting into the treatment group is unrelated to unobservable factors that affect the outcome variables. These estimators are not consistent otherwise. In order to estimate the extent to which such selection on unobservable may bias the estimates, we conducted a sensitive analysis with the bounding approach proposed by Rosenbaum (2002). The basic idea is to test whether the inference of the treatment effects may be altered by unobservable factors. It should be noted that, however, the method cannot inform us if there is hidden bias in the data. It only tells us how much of the hidden bias, if any, it would take to change inferences (Caliendo and Kopeinig, 2008).

In brief, the Rosenbaum bounds approach assumes that the participation probability π_i is not only determined by observed characteristicsX_i, but also by unobservable factorsu_i, so that:

$$\pi_i = \Pr(T_i = 1|X_i) = F(\beta X_i + \gamma u_i) \tag{7}$$

γ measures the effect of u_i on π_i. If no hidden bias exists, γ should be zero. If there is hidden bias, two individuals with the same propensity score calculated from the observed characteristics X_i would have different participation behaviors. Thus by varying the value of γ, one could assess the sensitivity of the results with respect to hidden bias and derive bounds of significance levels (Rosenbaum, 2002). But it is worth noting that this method only applies for one-to-one nearest neighbor and spline smoothing estimators.

Table 6 presents the values from Wilcoxon signed rank tests for the average treatment effect on the treated for spline smoothing estimator when setting the value of $\tau = e^{\gamma}$ at different levels[3]. First, we should describe how Table 6 should be interpreted. For each model and matching estimator, we increased the level of e^{γ}until the inference about the treatment effect is changed. We report the value of τ and the critical p-value. The bold cells in the table indicate that these appeared as statistically

[3] We used the *rbounds* module in Stata for continuous outcome variables (Gangl, 2004) and the *mhbounds* module for binary outcome variables (Becker and Caliendo, 2007)

significant when ATT's were estimated. For these cells, we report the value of τ for which the effect would become insignificant. For an ATT that was not statistically significant, we report critical value of τ at which degree the effect would become significant. We indicate the 5% level for estimates that turn from insignificant to significant and the 10% level for estimates that turn from significant to insignificant in the sense that these levels represent worst case scenarios (Drichoutis *et al.*, 2009).

Results in Table 6 tell us how strong the influence from unobserved factors should be to change the inference of our estimated ATTs. For example, in cell 1 vs. 0 for cognitive skills, the critical value for τ of 1.25 means that individuals with the same X differ in their odds of participation by a factor of 25%. The result states that the null hypothesis of no treatment effect would not be rejected if an unobserved variable caused the odds ratio of treatment assignment to differ between treatment and comparison groups by 1.25 and if this variable's effect on cognitive skills was so strong as to almost perfectly determine whether the cognitive skills would be higher for the treatment or the control case in each pair of matched cases in the data.

As shown in Table 6, in most cases, large value of τ are required to change the significance of our estimates, either from significant to insignificant, or from insignificant to significant. Thus, we can conclude that the statistical inference of our estimates would almost remain the same even if we had substantial unobserved heterogeneity. In other words, it is not likely that our main results of the effects of school transfers on child outcomes will change even in the presence of a large unobserved heterogeneity.

5. Robustness checks

Not all migrant children came to the county/district they currently live in after they started primary school. For those who came to the current county/district before starting primary school, their school transfer behaviors may be closer to local urban children as they did not experience any residential move after starting school. As Figure 3 shows, migrant children who came to their current county/district before age 6, the starting age of China's compulsory education, have a transfer pattern that is very similar to their local urban counterparts, while migrant children who came to their current county/district equal to or older than 6 years old have a much frequent school transfer pattern. Thus, we study the impact of school transfers on the outcome variables for those migrant children coming to the county/district they currently live in equal to or older than age 6.

Table 7 presents ATT for propensity scores matching estimations of children coming to the county/district they currently live in. For this subsample, the estimation results are consistent with our major results in terms of signs, but some of them lost statistical significance, which may partially be attributed to a small sample problem. A significant difference is that for migrant children who came to their current county/district equal to or older than age 6 and those who have transferred school just once have significantly higher cognitive skills than those who have never transferred.

We conduct another robustness check based on father's occupation characteristics. Most rural-to-urban migrant workers in China are low-skilled, and there may be difference in child outcomes between families where parents take high-skilled and low-skilled jobs. For comparison purposes, we divide our sample based on father's job characteristics. Figure 4 shows that the school transfer pattern for migrant children whose father take high-skill jobs and low-skilled jobs. Migrant children whose father take low-skill jobs or are unemployed are slightly more likely to transfer more than once. Table 8 exhibits the ATT for propensity scores matching estimations of children whose father takes low-skilled jobs or are unemployed, the results are very similar to the results of the full sample.

VI. Conclusions

Rural-to-urban migrant children have made up a fifth of children living in current urban China. However, compared with their urban local counterparts, migrant children are faced with many obstacles in pursuing personal development in cities. Unlike the institutional constraints that has drawn lots of attention from academia; the socioeconomically disadvantaged status of migrant children received little attention. In this paper, we study the impact of one special characteristic of migrant children that have been rarely discussed in literature, i.e., the high frequency of school transfers on health and academic outcomes of migrant children.

Using data from the 2013-2014 China Education Panel Survey, we employ Propensity Score Matching method to estimate the casual treatment effects. We depart from the binary treatment case since school transfers occur at different frequencies, which can be considered as different levels of treatment. Thus we not only compare the outcomes of migrant children who ever transferred schools to those who have never transferred, but also compare the outcomes of migrant children who transferred schools more times to those transferred fewer. We found that although school transfers in primary school years did not harm migrant children's cognitive skills, experiencing 3 or more school transfers significantly reduced their willingness to attend college in the future, and significantly increased their depression level. Experiencing school transfer ever has a substantial impact on the probability of experiencing grade retention, and it increases with number of school transfers. What is most interesting is that we find that compared to children who were never transferred to a school, transferring school just once has a statistically significantly positive impact on migrant children's cognitive skills and their willingness to attend college. It may correspond to that their parents take them to cities for better opportunities.

We examine the appropriateness of the CIA in our case by selecting a proper and exhausted set of conditional variables, and further conduct a sensitive test about how the inference of our estimated treatment effects would change if different scales of hidden bias exist. We also conducted robustness tests by excluding children who came to their current location before they turned 6 years old and whose father is high-skilled. Our results are robust to different specifications.

Reference

Abadie A., and G.W. Imbens. 2008. "On the failure of the bootstrap for matching estimators". *Econometrica* 76(6): 1537-57.

Alexander K.L., D. R. Entwisle, S.L. Dauber. 1996. "Children in motion: School transfers and elementary school performance". *Journal of Educational Research* 90(1): 3-12.

Astone N. M. and S.S. McLanahan. 1994. "Family structure, residential mobility, and school dropout: A research note". *Demography* 31(4): 575-84.

Becker S.O. and M. Caliendo, 2007. "Sensitivity analysis for average treatment effects" . *The Stata Journal* 7(1):71-83.

Bronfenbrenner U. 1979."Contexts of child rearing: Problems and prospects". *American Psychologist* 34(10): 844.

Burkam D.T., V.E. Lee and J. Dwyer. 2009. "School mobility in the early elementary grades: Frequency and impact from nationally-representative data." Paper prepared for the Workshop on the Impact of Mobility and Change on the Lives of Young Children, Schools, and Neighborhoods, June 29-30, 2009, Washington, DC. Available at: http://www.bocyf.org/children_who_move_burkam_paper.pdf, accessed March 1, 2017.

Caliendo M. and S. Kopeinig. 2008. "Some practical guidance for the implementation of propensity score matching". *Journal of Economic Surveys* 22(1): 31-72.

Cameron S.V. and J.J. Heckman. 1998. "Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts of American males". *Journal of Political Economy*106(2): 262-333.

Cameron S.V. and J.J. Heckman. 2001. "The dynamics of educational attainment for black, hispanic, and white males". *Journal of Political Economy* 109(3): 455-99.

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

Report of Migrant children in China, 2014. Available at http://www.ngocn.net/news/359422.html, accessed January 3, 2017.

DiPrete T.A. and M. Gangl. 2004. "Assessing bias in the estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables estimation with imperfect instruments". *Sociological Methodology* 34(1): 271-310.

Drichoutis A.C., R.M. Nayga and P. Lazaridis 2009. "Can nutritional label use influence body weight outcomes?" *Kyklos* 62(4): 500-25.

Ginsburg C, L.M. Richter and B. Fleisch .2011. "An analysis of associations between residential and school mobility and educational outcomes in South African urban children: The Birth to Twenty Cohort". *International Journal of Educational Development* 31(3): 213-22.

Gruman D.H., T.W. Harachi and R.D. Abbott.2008. "Longitudinal effects of student mobility on three dimensions of elementary school engagement". *Child development* 79(6): 1833-52.

Heckman J.J., H. Ichimura and P.E. Todd. 1997. "Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme". *The*

Review of Economic Studies 64(4): 605-54.

Heckman J.J., R.J. LaLonde and J.A. Smith. 1999. "The economics and econometrics of active labor market programs". *Handbook of Labor Economics*: 1865-2097.

Heckman J.J. and S. Navarro-Lozano. 2004. "Using matching, instrumental variables, and control functions to estimate economic choice models". *Review of Economics and Statistics* 86(1): 30-57.

Heckman J.J., J. Stixrud and S. Urzua. 2006. "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior". *Journal of Labor economics* 24(3): 411-82.

Jacob B.A. and L. Lefgren. 2009. "The effect of grade retention on high school completion". *American Economic Journal: Applied Economics* 1(3): 33-58.

Lai F, C. Liu, R. Luo. 2014. "The education of China's migrant children: The missing link in China's education system". *International Journal of Educational Development* 37: 68-77.

Lechner M. 2008. "A note on endogenous control variables in causal studies". *Statistics & Probability Letters* 78(2): 190-5.

Lechner M. 2002. "Program heterogeneity and propensity score matching: An application to the evaluation of active labor market policies". *Review of Economics and Statistics* 84(2): 205-220.

Manacorda M. 2012. "The cost of grade retention". *Review of Economics and Statistics* 94(2): 596-606.

Mehana M. and A.J. Reynolds. 2004. "School mobility and achievement: A meta-analysis". *Children and Youth Services Review* 26(1): 93-119.

Pribesh S. and D.B. Downey. 1999. "Why are residential and school moves associated with poor school performance?". *Demography* 36(4): 521-534.

Rumberger R.W. and K.A. Larson. 1998. "Student mobility and the increased risk of high school dropout". *American Journal of Education* 107(1): 1-35.

Rosenbaum P.R. 2002. *Observational Studies*. Springer New York.

Rubin D.B. 1991. "Practical implications of modes of statistical inference for causal effects and the critical role of the assignment mechanism". *Biometrics*: 1213-34.

Strand S, Demie F. 2006. "Pupil mobility, attainment and progress in primary school". *British Educational Research Journal* 32(4): 551-68.

South S.J., D.L. Haynie and S. Bose. 2007. "Student mobility and school dropout". *Social Science Research* 36(1): 68-94.

Tafreschi D. and P. Thiemann. 2016. "Doing it twice, getting it right? The effects of grade retention and course repetition in higher education". *Economics of Education Review* 55: 198-219.

Temple J.A.and A.J. Reynolds. 2000. "School mobility and achievement: Longitudinal findings from an urban cohort". *Journal of School Psychology* 37(4): 355-77.

Xie J., X. Niu and Y. Xie. 2011. "Investigation on the Educational Problems of the Along-with Offspring of Rural Migrant Workers in Cities". *Chinese Journal of Population Science* 1: 1-10.

Xu Z., J. Hannaway, and S. D'Souza. 2009. Student Transience in North Carolina: The Effect of School Mobility on Student Outcomes Using Longitudinal Data. *Working Paper 22*, National Center for Analysis of Longitudinal Data in Education Research.

Zeng S. and Q. Li. 2007. The Research on Migrant Children's Social Adaptation: Its Current Situation, Problems and Solutions. *Psychological Science-Shanghai* 30(6): 1426.

Table 1. Variable description

	Transfers=0		Transfers =1		Transfers =2		Transfers >=3	
	Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd
Outcome variables								
Cognitive skills	-0.03	0.84	0.06	0.82	-0.08	0.77	-0.11	0.83
Grade retention ever (yes = 1)	0.09	0.29	0.24	0.43	0.36	0.48	0.51	0.50
Willingness to attend college (yes = 1)	0.80	0.40	0.84	0.36	0.78	0.42	0.63	0.48
Depression level	-0.11	0.98	-0.19	0.86	0.04	0.90	0.30	1.12
Child variables								
Child's age	13.10	0.80	13.30	0.93	13.50	0.91	13.50	0.88
Child's age when starting primary school	6.52	1.04	6.55	1.00	6.58	1.10	6.53	1.00
Boy = 1	0.52	0.50	0.53	0.50	0.51	0.50	0.56	0.50
Number of siblings	0.91	0.82	0.95	0.74	0.92	0.72	1.15	0.70
Ethnics: *Han* = 1	0.95	0.21	0.91	0.29	0.90	0.31	0.93	0.25
Ever attended a kindergarten = 1	0.84	0.36	0.79	0.41	0.73	0.45	0.64	0.48
Ever in bad disease before primary school = 1	0.07	0.26	0.06	0.24	0.05	0.22	0.08	0.27
Migrating across province = 1	0.55	0.50	0.65	0.48	0.64	0.48	0.71	0.46
Family income variables								
Mother's occupation: *High-skill*[1]=1	0.06	0.23	0.06	0.24	0.05	0.22	0.01	0.09
Low-skill=1	0.72	0.45	0.79	0.41	0.77	0.43	0.74	0.44
Unemployed or other = 1	0.22	0.42	0.15	0.36	0.18	0.39	0.26	0.44
Father's occupation: *High-skill*[1]=1	0.14	0.35	0.14	0.34	0.13	0.34	0.08	0.27
Low-skill=1	0.76	0.43	0.81	0.40	0.78	0.42	0.82	0.39
Unemployed or other = 1	0.10	0.30	0.06	0.23	0.09	0.29	0.10	0.30
Parent-report family income before primary school: *Low*=1	0.25	0.43	0.34	0.47	0.25	0.43	0.37	0.48
Middle = 1	0.72	0.45	0.64	0.48	0.73	0.45	0.61	0.49
High = 1	0.04	0.19	0.02	0.15	0.03	0.16	0.03	0.16
Currently receiving the minimum Subsistence allowance = 1	0.04	0.19	0.06	0.24	0.10	0.31	0.08	0.27
Family education motivation variables								
Mother's education: *Primary school* = 1	0.29	0.45	0.32	0.47	0.32	0.47	0.40	0.49
Middle school = 1	0.58	0.49	0.54	0.50	0.52	0.50	0.48	0.50
High school = 1	0.11	0.31	0.12	0.33	0.08	0.27	0.08	0.27
>High school = 1	0.03	0.16	0.02	0.14	0.08	0.27	0.04	0.20
Father's education: *Primary school* = 1	0.16	0.37	0.16	0.37	0.21	0.41	0.21	0.41
Middle school = 1	0.58	0.49	0.60	0.49	0.55	0.50	0.62	0.49
High school = 1	0.22	0.42	0.19	0.39	0.16	0.37	0.13	0.34
>High school = 1	0.04	0.20	0.05	0.21	0.09	0.29	0.03	0.18
Parental relationship: *Good*=1	0.84	0.37	0.83	0.38	0.83	0.38	0.75	0.43
Father often gets drunk=1	0.07	0.25	0.08	0.27	0.09	0.29	0.14	0.35
Parental strictness on the child	0.13	0.95	0.15	0.91	-0.14	1.13	0.11	1.09

Note: High-skill occupations include: Government official, staff of public institutions, civil servant; Middle/Senior management personnel of enterprises/corporations; Teacher, engineer, doctor, lawyer. Low-skill categories includes: Technical worker (including driver); Ordinary staff or worker in business or service industry; Self-employed worker; Peasant.

Table 2. Probit model for school transfer

VARIABLES	(1) >=1 vs. 0	(2) 1 vs. 0	(3) 2 vs. 0	(4) >=3 vs. 0	(5) 2 vs. 1	(6) >=3 vs. 1	(7) >=3 vs. 2
Child variables							
Child's age	0.214***	0.138**	0.310***	0.303***	0.147*	0.135*	-0.019
	(0.050)	(0.058)	(0.087)	(0.078)	(0.088)	(0.081)	(0.111)
Child's age when starting	0.006	0.007	0.027	-0.033	0.015	-0.043	-0.093
primary school	(0.042)	(0.048)	(0.070)	(0.062)	(0.080)	(0.072)	(0.101)
Boy	-0.030	-0.033	-0.121	0.051	-0.065	0.119	0.184
	(0.086)	(0.098)	(0.144)	(0.134)	(0.166)	(0.151)	(0.208)
Number of siblings	0.045	0.012	-0.020	0.154*	-0.038	0.202**	0.279*
	(0.055)	(0.063)	(0.091)	(0.080)	(0.112)	(0.102)	(0.143)
Ethnics: *Han*	-0.351**	-0.399**	-0.365	-0.188	-0.002	0.249	0.319
	(0.169)	(0.190)	(0.269)	(0.271)	(0.276)	(0.270)	(0.368)
Ever attended a kindergarten	-0.323***	-0.186	-0.362**	-0.488***	-0.178	-0.326**	-0.176
	(0.105)	(0.125)	(0.173)	(0.147)	(0.190)	(0.163)	(0.218)
Ever in bad disease before	-0.058	-0.068	-0.232	0.067	-0.201	0.095	0.276
primary school	(0.168)	(0.197)	(0.301)	(0.240)	(0.346)	(0.284)	(0.422)
Migrating across province	0.308***	0.269***	0.223	0.343**	-0.006	0.111	0.148
	(0.087)	(0.099)	(0.146)	(0.133)	(0.170)	(0.159)	(0.215)
Family income variables							
Mother's occupation: *high-*	0.043	0.303	0.064	-0.957*	-0.330	-1.217**	-1.296*
skill	(0.223)	(0.244)	(0.392)	(0.492)	(0.436)	(0.532)	(0.756)
Low-skill	0.075	0.155	0.190	-0.203	0.021	-0.350*	-0.343
	(0.116)	(0.136)	(0.206)	(0.172)	(0.242)	(0.196)	(0.283)
Father's occupation:	0.212	0.368	-0.072	0.096	-0.244	-0.370	0.030
high-skill	(0.203)	(0.238)	(0.344)	(0.312)	(0.400)	(0.377)	(0.490)
Low-skill	0.233	0.335*	-0.008	0.204	-0.241	-0.187	0.161
	(0.166)	(0.202)	(0.275)	(0.239)	(0.335)	(0.301)	(0.387)
Parent-report family income	0.338	0.482	-0.001	0.114	-0.370	-0.113	0.438
before primary school: *low*	(0.254)	(0.296)	(0.423)	(0.399)	(0.515)	(0.475)	(0.660)
Middle	0.119	0.231	-0.029	-0.140	-0.126	-0.141	0.047
	(0.246)	(0.286)	(0.403)	(0.389)	(0.504)	(0.469)	(0.646)
Currently receiving the	0.449**	0.318	0.518*	0.494*	0.343	0.355	-0.041
minimum Subsistence	(0.186)	(0.221)	(0.283)	(0.266)	(0.313)	(0.294)	(0.359)
allowance							
Family education motivation variables							
Mother's education: *Primary*	-0.294	0.185	-0.726*	-0.834**	-1.022**	-1.173**	-0.086
school	(0.282)	(0.369)	(0.413)	(0.397)	(0.519)	(0.500)	(0.555)
Middle school	-0.319	0.160	-0.711*	-0.855**	-0.985*	-1.200**	-0.096
	(0.273)	(0.360)	(0.393)	(0.381)	(0.511)	(0.490)	(0.533)
High school	-0.096	0.406	-0.649	-0.705*	-1.091**	-1.209**	0.046
	(0.286)	(0.371)	(0.427)	(0.418)	(0.528)	(0.523)	(0.647)
Mother's education: *Primary*	-0.071	-0.111	-0.138	0.161	0.057	0.198	0.167

school	(0.246)	(0.280)	(0.390)	(0.410)	(0.449)	(0.456)	(0.613)
Middle school	-0.012	-0.005	-0.195	0.266	-0.154	0.132	0.233
	(0.227)	(0.257)	(0.362)	(0.383)	(0.419)	(0.426)	(0.580)
High school	-0.208	-0.170	-0.362	-0.069	-0.194	-0.007	0.120
	(0.229)	(0.260)	(0.363)	(0.384)	(0.429)	(0.437)	(0.577)
Parental relationship	-0.057	-0.006	0.190	-0.228	0.169	-0.275	-0.632**
	(0.111)	(0.129)	(0.200)	(0.156)	(0.227)	(0.181)	(0.272)
Father often gets drunk	0.170	0.090	0.119	0.313	0.032	0.179	0.060
	(0.154)	(0.182)	(0.259)	(0.216)	(0.301)	(0.240)	(0.327)
Parental strictness	-0.028	-0.008	-0.152**	0.016	-0.183**	0.042	0.186*
	(0.043)	(0.051)	(0.070)	(0.064)	(0.083)	(0.074)	(0.096)
Observations	989	795	602	642	347	387	194

Note: Standard error in parenthesis; ***, **, * statistically significant at the 1%, 5%, 10% level.

Table 3. Average treatment effects on the treated (ATT) for propensity scores matching estimations.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	>=1 vs. 0[1]	1 vs. 0[1]	2 vs. 0[1]	>=3 vs. 0[1]	2 vs. 1[1]	>=3 vs. 1[1]	>=3 vs. 2[1]
Cognitive skills							
Unmatched	0.018	0.083	-0.054	-0.087	-0.138	-0.170*[3]	-0.032
	(0.053)	(0.062)	(0.101)	(0.085)	(0.105)	(0.091)	(0.118)
Nearest neighborhood[2]	0.111	0.281***	0.321**	0.067	-0.197	-0.077	0.078
	(0.075)	(0.090)	(0.146)	(0.129)	(0.155)	(0.133)	(0.182)
Local linear regression	0.088	0.139**	0.042	0.010	-0.142	-0.055	-0.167
	(0.060)	(0.067)	(0.109)	(0.102)	(0.115)	(0.114)	(0.164)
Spline-smoothing	0.087	0.147**	0.049	0.026	-0.133	-0.060	-0.144
	(0.061)	(0.068)	(0.106)	(0.099)	(0.109)	(0.110)	(0.166)
Kernel	0.093	0.143**	0.028	0.023	-0.138	-0.041	-0.156
	(0.060)	(0.067)	(0.110)	(0.104)	(0.117)	(0.115)	(0.165)
Radius, Caliper=0.1	0.082	0.131**	0.018	0.015	-0.129	-0.073	-0.118
	(0.059)	(0.065)	(0.103)	(0.100)	(0.110)	(0.110)	(0.165)
Radius, Caliper=0.01	0.122*	0.127*	0.048	0.056	-0.189	-0.142	0.031
	(0.064)	(0.073)	(0.136)	(0.123)	(0.129)	(0.130)	(0.201)
Grade retention							
Unmatched	0.242***	0.155***	0.274***	0.423***	0.119**	0.268***	0.149**
	(0.024)	(0.026)	(0.039)	(0.034)	(0.057)	(0.050)	(0.073)
Nearest neighborhood[2]	0.207***	0.093**	0.171**	0.328***	0.013	0.207***	0.063
	(0.032)	(0.040)	(0.077)	(0.065)	(0.087)	(0.077)	(0.112)
Local linear regression	0.203***	0.134***	0.229***	0.375***	0.054	0.191***	0.122
	(0.029)	(0.030)	(0.060)	(0.054)	(0.072)	(0.062)	(0.097)
Spline-smoothing	0.206***	0.135***	0.222***	0.367***	0.072	0.179***	0.135
	(0.028)	(0.029)	(0.060)	(0.053)	(0.065)	(0.062)	(0.090)
Kernel	0.205***	0.137***	0.227***	0.367***	0.045	0.197***	0.100
	(0.029)	(0.030)	(0.064)	(0.055)	(0.076)	(0.065)	(0.103)
Radius, Caliper=0.1	0.212***	0.140***	0.241***	0.374***	0.066	0.201***	0.134
	(0.028)	(0.028)	(0.059)	(0.053)	(0.066)	(0.060)	(0.089)
Radius, Caliper=0.01	0.204***	0.140***	0.196***	0.364***	0.095	0.204***	0.068
	(0.030)	(0.033)	(0.066)	(0.062)	(0.082)	(0.075)	(0.123)
Willingness to attend college							
Unmatched	-0.022	0.043	-0.023	-0.169***	-0.065	-0.212***	-0.147**
	(0.026)	(0.029)	(0.049)	(0.042)	(0.049)	(0.045)	(0.067)
Nearest neighborhood[2]	-0.007	0.156***	-0.039	-0.129*	-0.065	-0.164**	-0.171**
	(0.036)	(0.046)	(0.073)	(0.069)	(0.072)	(0.071)	(0.087)
Local linear regression	0.010	0.072**	-0.003	-0.109**	-0.094	-0.172***	-0.194**
	(0.029)	(0.033)	(0.058)	(0.055)	(0.063)	(0.058)	(0.081)
Spline-smoothing	0.011	0.072**	0.008	-0.118**	-0.086	-0.172***	-0.182**
	(0.029)	(0.033)	(0.058)	(0.054)	(0.063)	(0.058)	(0.078)
Kernel	0.010	0.070**	0.001	-0.109**	-0.091	-0.172***	-0.182**

	(0.029)	(0.032)	(0.058)	(0.056)	(0.063)	(0.059)	(0.088)
Radius, Caliper=0.1	0.006	0.063**	-0.004	-0.127**	-0.085	-0.177***	-0.175**
	(0.030)	(0.030)	(0.055)	(0.055)	(0.059)	(0.057)	(0.081)
Radius, Caliper=0.01	0.002	0.084**	0.026	-0.127**	-0.038	-0.197***	-0.188*
	(0.033)	(0.035)	(0.064)	(0.063)	(0.068)	(0.069)	(0.105)
Depression level							
Unmatched	0.089	-0.073	0.157	0.417***	0.231**	0.490***	0.260*
	(0.062)	(0.071)	(0.118)	(0.103)	(0.112)	(0.105)	(0.153)
Nearest neighborhood[2]	0.055	-0.066	0.255	0.477***	0.212	0.305*	0.161
	(0.087)	(0.102)	(0.170)	(0.166)	(0.156)	(0.173)	(0.224)
Local linear regression	0.044	-0.115	0.173	0.290**	0.248*	0.416***	0.138
	(0.073)	(0.080)	(0.127)	(0.143)	(0.131)	(0.128)	(0.191)
Spline-smoothing	0.048	-0.099	0.149	0.302**	0.226*	0.425***	0.148
	(0.071)	(0.078)	(0.128)	(0.134)	(0.130)	(0.128)	(0.189)
Kernel	0.052	-0.092	0.168	0.322**	0.237*	0.418***	0.164
	(0.072)	(0.078)	(0.131)	(0.145)	(0.133)	(0.131)	(0.201)
Radius, Caliper=0.1	0.057	-0.093	0.150	0.338**	0.228*	0.424***	0.170
	(0.069)	(0.069)	(0.125)	(0.137)	(0.128)	(0.126)	(0.194)
Radius, Caliper=0.01	0.060	-0.058	0.175	0.384***	0.189	0.443***	0.244
	(0.075)	(0.086)	(0.148)	(0.148)	(0.157)	(0.158)	(0.251)
Observations	989	795	602	642	347	387	194

Note: 1. Bootstrap standard errors for ATT except nearest neighbor, N = 800 replications.

2. With replacement, no caliper.

3. ***, **, * statistically significant at the 1%, 5%, 10% level.

Table 4. Number of treated individuals lost due to common support requirement and range of the propensity scores after common support imposition

	(1) >=1 vs. 0	(2) 1 vs. 0	(3) 2 vs. 0	(4) >=3 vs. 0	(5) 2 vs. 1	(6) >=3 vs. 1	(7) >=3 vs. 2
Before	989	795	602	642	347	387	194
After	984	795	601	641	347	386	188
Lost in %	0.51	0.00	0.17	0.16	0.00	0.26	3.09

Table 5. Quality of matching indicators

		(1) >=1 vs. 0	(2) 1 vs. 0	(3) 2 vs. 0	(4) >=3 vs. 0	(5) 2 vs. 1	(6) >=3 vs. 1	(7) >=3 vs. 2
	Mean absolute bias	10.55	9.079	11.93	17.03	9.846	13.48	13.60
Unmatched	Pseudo R2	0.060	0.043	0.095	0.134	0.065	0.088	0.097
	P-value of LR chi2	0.000	0.007	0.008	0.000	0.478	0.013	0.389
Nearest Neighbor/	Mean absolute bias	5.148	4.968	8.433	9.494	9.796	8.895	10.33
Local Linear/	Pseudo R2	0.016	0.014	0.041	0.076	0.054	0.050	0.054
Spline-smoothing	P-value of LR chi2	0.687	0.993	0.997	0.402	0.985	0.888	0.844
Kernel	Mean absolute bias	1.162	1.151	4.102	2.767	4.773	3.793	4.540
	Pseudo R2	0.001	0.001	0.009	0.006	0.013	0.012	0.015
	P-value of LR chi2	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Radius, cal=0.1	Mean absolute bias	1.440	2.025	3.671	3.645	3.359	2.943	4.036
	Pseudo R2	0.002	0.003	0.012	0.008	0.009	0.007	0.011
	P-value of LR chi2	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Radius, cal=0.01	Mean absolute bias	1.838	2.136	3.231	3.259	7.088	4.563	8.172
	Pseudo R2	0.002	0.003	0.009	0.010	0.030	0.016	0.039
	P-value of LR chi2	1.000	1.000	1.000	1.000	1.000	1.000	0.997

Table 6. Rosenbaum bounds for treatment effects

	>=1 vs. 0	1 vs. 0	2 vs. 0	>=3 vs. 0	2 vs. 1	>=3 vs. 1	>=3 vs. 2
Cognitive skills	1.01(0.020) 1.10(0.104)	**1.25(0.109)**	1.49(0.048)	1.33(0.047)	1.01(0.047)	1.23(0.047)	1.01(0.048) 1.10(0.101)
Grade retention	**3.06(0.104)**	**1.40(0.101)**	**1.28(0.101)**	**3.21(0.100)**	1.66(0.049)	**1.66(0.102)**	1.12(0.049)
Education expectation	1.51(0.049)	**1.68(0.100)**	2.19(0.049)	**1.20(0.101)**	1.66(0.049)	**1.33(0.102)**	**1.79(0.100)**
Depression level	1.24(0.048)	1.01(0.005) 1.22(0.105)	1.16(0.048)	**1.18(0.102)**	**1.15(0.106)**	**1.59(0.102)**	1.28(0.047)

Note: Critical P-value in parentheses.

Table 7. ATT for propensity scores matching estimations of children coming the county/district they currently live in equal to or older than age 6.

VARIABLES	(1) >=1 vs. 0[1]	(2) 1 vs. 0[1]	(3) 2 vs. 0[1]	(4) >=3 vs. 0[1]	(5) 2 vs. 1[1]	(6) >=3 vs. 1[1]	(7) >=3 vs. 2[1]
Cognitive skills							
Unmatched	0.082	0.145	0.040	-0.031	-0.105	-0.150	-0.014
	(0.082)	(0.094)	(0.143)	(0.117)	(0.129)	(0.108)	(0.148)
Nearest neighborhood[2]	0.145	0.209	-0.046	-0.162	-0.154	-0.175	-0.171
	(0.106)	(0.136)	(0.230)	(0.181)	(0.185)	(0.167)	(0.215)
Local linear regression	0.106	0.166	0.025	0.019	-0.092	-0.186	-0.270
	(0.090)	(0.105)	(0.196)	(0.158)	(0.168)	(0.147)	(0.262)
Spline-smoothing	0.097	0.192*	0.033	0.054	-0.104	-0.209	-0.210
	(0.093)	(0.101)	(0.182)	(0.152)	(0.157)	(0.141)	(0.241)
Kernel	0.118	0.177*	-0.010	0.035	-0.146	-0.158	-0.252
	(0.098)	(0.106)	(0.215)	(0.158)	(0.175)	(0.159)	(0.259)
Radius, Caliper=0.1	0.110	0.197*	0.023	0.048	-0.063	-0.174	-0.096
	(0.092)	(0.105)	(0.186)	(0.143)	(0.156)	(0.145)	(0.249)
Radius, Caliper=0.01	0.163	0.172	-0.111	-0.045	-0.035	-0.212	-0.057
	(0.109)	(0.130)	(0.257)	(0.195)	(0.228)	(0.187)	(0.343)
Grade retention							
Unmatched	0.223***	0.138***	0.213***	0.398***	0.075	0.254***	0.176*
	(0.045)	(0.047)	(0.066)	(0.059)	(0.077)	(0.066)	(0.093)
Nearest neighborhood[2]	0.148***	0.106	0.158	0.338***	-0.065	0.203**	0.000
	(0.054)	(0.063)	(0.108)	(0.091)	(0.112)	(0.099)	(0.137)
Local linear regression	0.201***	0.132**	0.190**	0.361***	-0.028	0.163*	0.019
	(0.049)	(0.054)	(0.096)	(0.085)	(0.105)	(0.094)	(0.151)
Spline-smoothing	0.201***	0.129**	0.189**	0.370***	-0.003	0.138	0.031
	(0.047)	(0.051)	(0.092)	(0.078)	(0.097)	(0.088)	(0.141)
Kernel	0.200***	0.123**	0.194*	0.355***	-0.048	0.172*	0.008
	(0.049)	(0.054)	(0.099)	(0.085)	(0.114)	(0.099)	(0.159)
Radius, Caliper=0.1	0.205***	0.120**	0.186**	0.362***	-0.025	0.176**	0.062
	(0.047)	(0.053)	(0.090)	(0.082)	(0.101)	(0.086)	(0.144)
Radius, Caliper=0.01	0.221***	0.138***	0.206**	0.398***	0.048	0.248***	0.135
	(0.042)	(0.047)	(0.082)	(0.069)	(0.083)	(0.072)	(0.108)
Willingness to attend college							
Unmatched	-0.025	0.046	-0.036	-0.156**	-0.081	-0.207***	-0.102
	(0.042)	(0.044)	(0.069)	(0.061)	(0.064)	(0.058)	(0.089)
Nearest neighborhood[2]	-0.021	0.131**	0.053	-0.163*	-0.109	-0.076	-0.217*
	(0.055)	(0.068)	(0.108)	(0.093)	(0.087)	(0.093)	(0.124)
Local linear regression	-0.014	0.059	-0.002	-0.129	-0.102	-0.118	-0.219*
	(0.052)	(0.057)	(0.095)	(0.092)	(0.092)	(0.084)	(0.126)
Spline-smoothing	-0.022	0.052	-0.006	-0.118	-0.101	-0.135	-0.205*
	(0.048)	(0.053)	(0.087)	(0.086)	(0.085)	(0.082)	(0.121)
Kernel	-0.031	0.062	-0.002	-0.140	-0.099	-0.099	-0.215*

	(0.049)	(0.059)	(0.103)	(0.090)	(0.089)	(0.086)	(0.130)
Radius, Caliper=0.1	-0.032	0.052	-0.004	-0.133	-0.083	-0.116	-0.193
	(0.047)	(0.050)	(0.090)	(0.086)	(0.091)	(0.082)	(0.120)
Radius, Caliper=0.01	-0.029	0.046	-0.023	-0.156**	-0.070	-0.199***	-0.119
	(0.042)	(0.043)	(0.074)	(0.068)	(0.075)	(0.069)	(0.107)
Depression level							
Unmatched	0.017	-0.177	0.148	0.289*	0.324**	0.494***	0.117
	(0.099)	(0.109)	(0.171)	(0.148)	(0.145)	(0.130)	(0.191)
Nearest neighborhood[2]	-0.054	-0.314*	0.205	0.191	0.363*	0.498***	0.293
	(0.141)	(0.160)	(0.233)	(0.212)	(0.205)	(0.186)	(0.243)
Local linear regression	-0.090	-0.230*	0.112	0.145	0.302*	0.357*	0.168
	(0.121)	(0.130)	(0.210)	(0.196)	(0.182)	(0.183)	(0.298)
Spline-smoothing	-0.047	-0.227*	0.122	0.104	0.301*	0.362**	0.135
	(0.115)	(0.133)	(0.201)	(0.182)	(0.168)	(0.165)	(0.262)
Kernel	-0.029	-0.231	0.143	0.164	0.300*	0.376**	0.166
	(0.116)	(0.141)	(0.218)	(0.191)	(0.182)	(0.188)	(0.295)
Radius, Caliper=0.1	-0.008	-0.230*	0.107	0.147	0.323*	0.370**	0.151
	(0.106)	(0.132)	(0.214)	(0.183)	(0.184)	(0.165)	(0.277)
Radius, Caliper=0.01	0.017	-0.177	0.080	0.289**	0.361**	0.477***	0.096
	(0.104)	(0.111)	(0.177)	(0.147)	(0.160)	(0.153)	(0.226)
Observations	442	314	202	226	208	229	124

Note: 1. Bootstrap standard errors for ATT except nearest neighbor, N = 800 replications.

2. With replacement, no caliper.

3. ***, **, * statistically significant at the 1%, 5%, 10% level.

Table 8. ATT for propensity scores matching estimations of children whose father takes low-skill jobs

VARIABLES	(1) >=1 vs. 0[1]	(2) 1 vs. 0[1]	(3) 2 vs. 0[1]	(4) >=3 vs. 0[1]	(5) 2 vs. 1[1]	(6) >=3 vs. 1[1]	(7) >=3 vs. 2[1]
Cognitive skills							
Unmatched	-0.009	0.059	-0.105	-0.093	-0.164	-0.143	0.025
	(0.056)	(0.066)	(0.108)	(0.088)	(0.112)	(0.095)	(0.126)
Nearest neighborhood[2]	-0.015	0.120	-0.080	0.002	-0.021	-0.173	-0.120
	(0.082)	(0.096)	(0.145)	(0.131)	(0.153)	(0.146)	(0.174)
Local linear regression	0.058	0.097	-0.001	0.020	-0.133	-0.076	-0.090
	(0.066)	(0.072)	(0.122)	(0.102)	(0.128)	(0.119)	(0.179)
Spline-smoothing	0.060	0.107	-0.004	0.034	-0.132	-0.093	-0.084
	(0.063)	(0.074)	(0.118)	(0.098)	(0.116)	(0.110)	(0.168)
Kernel	0.059	0.104	-0.010	0.017	-0.130	-0.108	-0.101
	(0.060)	(0.070)	(0.127)	(0.111)	(0.127)	(0.122)	(0.188)
Radius, Caliper=0.1	0.055	0.102	-0.026	0.008	-0.126	-0.078	-0.077
	(0.060)	(0.068)	(0.121)	(0.106)	(0.122)	(0.113)	(0.162)
Radius, Caliper=0.01	0.070	0.110	-0.036	0.036	-0.039	-0.144	0.008
	(0.067)	(0.078)	(0.147)	(0.126)	(0.152)	(0.146)	(0.226)
Grade retention							
Unmatched	0.240***	0.139***	0.295***	0.424***	0.156**	0.281***	0.140*
	(0.026)	(0.028)	(0.042)	(0.037)	(0.061)	(0.053)	(0.078)
Nearest neighborhood[2]	0.214***	0.060	0.246***	0.458***	0.045	0.222***	0.074
	(0.036)	(0.044)	(0.078)	(0.057)	(0.093)	(0.086)	(0.114)
Local linear regression	0.202***	0.119***	0.256***	0.381***	0.084	0.222***	0.106
	(0.030)	(0.032)	(0.069)	(0.056)	(0.076)	(0.065)	(0.096)
Spline-smoothing	0.206***	0.120***	0.248***	0.377***	0.082	0.197***	0.100
	(0.029)	(0.032)	(0.063)	(0.053)	(0.072)	(0.064)	(0.099)
Kernel	0.205***	0.119***	0.251***	0.385***	0.087	0.231***	0.097
	(0.030)	(0.031)	(0.068)	(0.061)	(0.076)	(0.068)	(0.098)
Radius, Caliper=0.1	0.211***	0.128***	0.258***	0.388***	0.100	0.219***	0.086
	(0.028)	(0.031)	(0.064)	(0.060)	(0.074)	(0.065)	(0.092)
Radius, Caliper=0.01	0.211***	0.095***	0.252***	0.398***	0.026	0.219***	0.117
	(0.032)	(0.036)	(0.075)	(0.068)	(0.090)	(0.078)	(0.135)
Willingness to attend college							
Unmatched	-0.012	0.065**	-0.043	-0.159***	-0.108**	-0.229***	-0.113
	(0.028)	(0.032)	(0.054)	(0.046)	(0.052)	(0.046)	(0.073)
Nearest neighborhood[2]	0.010	0.125***	0.000	-0.131*	-0.149**	-0.269***	-0.221**
	(0.041)	(0.048)	(0.080)	(0.069)	(0.067)	(0.069)	(0.103)
Local linear regression	0.025	0.101***	-0.002	-0.118*	-0.132**	-0.223***	-0.133
	(0.033)	(0.035)	(0.065)	(0.063)	(0.065)	(0.061)	(0.091)
Spline-smoothing	0.024	0.098***	0.007	-0.110*	-0.134**	-0.206***	-0.131
	(0.033)	(0.034)	(0.063)	(0.058)	(0.064)	(0.059)	(0.088)
Kernel	0.022	0.105***	0.008	-0.111*	-0.120*	-0.229***	-0.108

	(0.031)	(0.035)	(0.065)	(0.062)	(0.063)	(0.064)	(0.099)
Radius, Caliper=0.1	0.018	0.092***	0.004	-0.118**	-0.121*	-0.216***	-0.125
	(0.033)	(0.033)	(0.062)	(0.056)	(0.064)	(0.059)	(0.086)
Radius, Caliper=0.01	0.026	0.130***	-0.006	-0.124*	-0.057	-0.224***	-0.123
	(0.036)	(0.039)	(0.073)	(0.066)	(0.074)	(0.073)	(0.128)
Depression level							
Unmatched	0.017	-0.177	0.148	0.289*	0.324**	0.494***	0.117
	(0.099)	(0.109)	(0.171)	(0.148)	(0.145)	(0.130)	(0.191)
Nearest neighborhood[2]	-0.084	-0.017	0.165	0.220	0.313**	0.282	-0.024
	(0.100)	(0.109)	(0.167)	(0.166)	(0.170)	(0.179)	(0.221)
Local linear regression	0.055	-0.102	0.212	0.327**	0.269*	0.431***	0.038
	(0.084)	(0.088)	(0.143)	(0.143)	(0.142)	(0.138)	(0.194)
Spline-smoothing	0.056	-0.091	0.208	0.267*	0.267**	0.428***	0.052
	(0.077)	(0.085)	(0.127)	(0.141)	(0.136)	(0.132)	(0.211)
Kernel	0.070	-0.061	0.197	0.274*	0.256*	0.427***	0.044
	(0.076)	(0.088)	(0.144)	(0.143)	(0.152)	(0.136)	(0.205)
Radius, Caliper=0.1	0.068	-0.090	0.192	0.289**	0.258*	0.438***	0.073
	(0.075)	(0.081)	(0.137)	(0.134)	(0.138)	(0.134)	(0.200)
Radius, Caliper=0.01	0.069	-0.038	0.182	0.357**	0.287	0.354**	0.260
	(0.084)	(0.092)	(0.163)	(0.156)	(0.174)	(0.162)	(0.285)
Observations	859	684	518	553	300	335	174

Note: 1. Bootstrap standard errors for ATT except nearest neighbor, N = 800 replications.

2. With replacement, no caliper.

3. ***, **, * statistically significant at the 1%, 5%, 10% level.

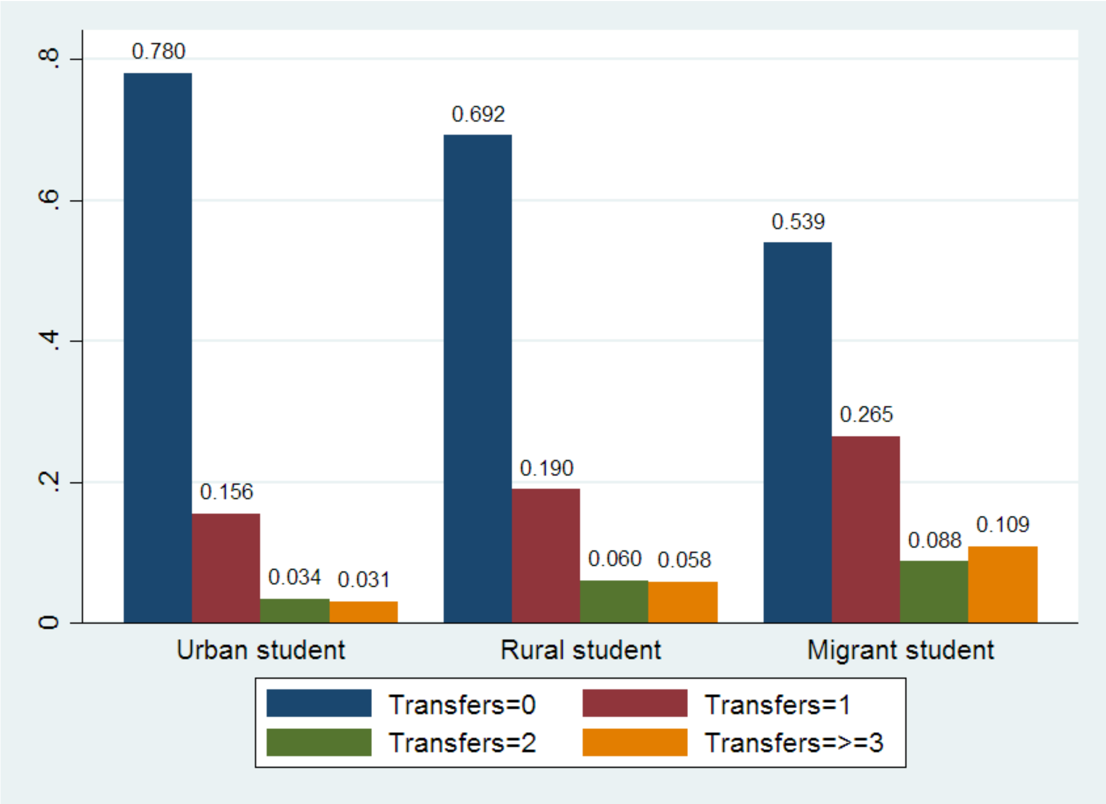

Figure 1. Migrant types and school transfers of children (Source: 2013-2014 China Education Panel Survey, CEPS).

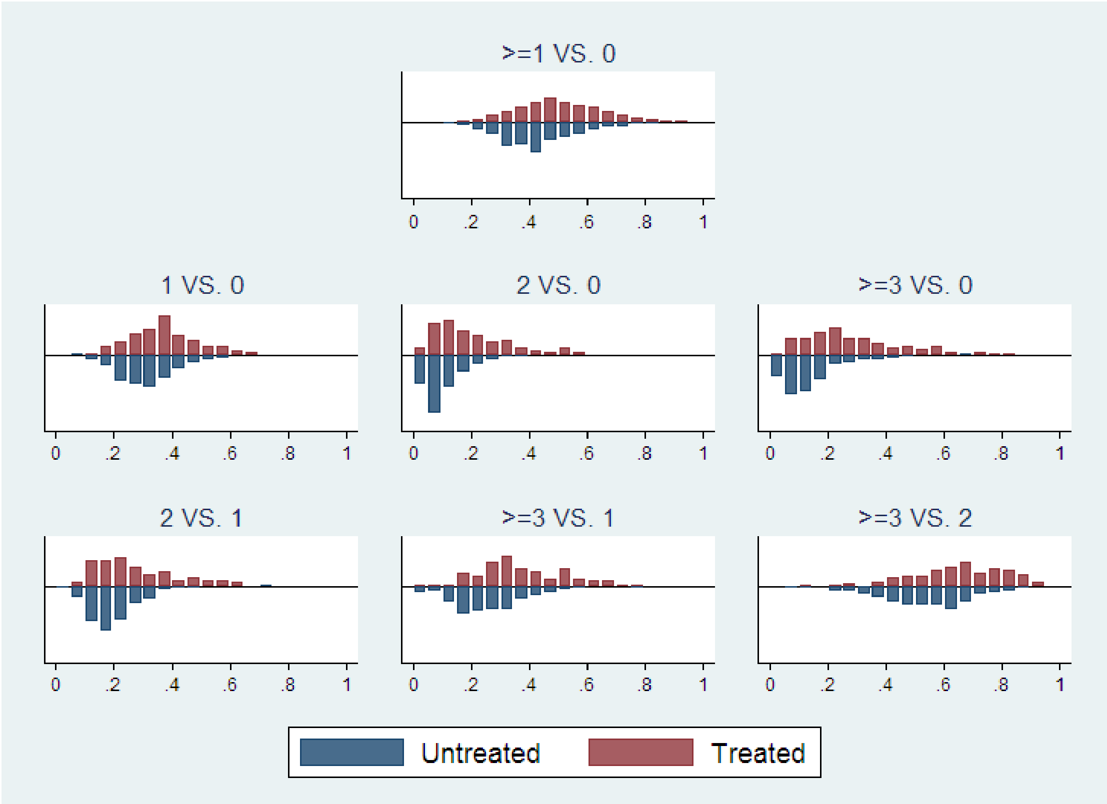

Figure 2. Propensity scores (frequencies for probability intervals by treatments and models)

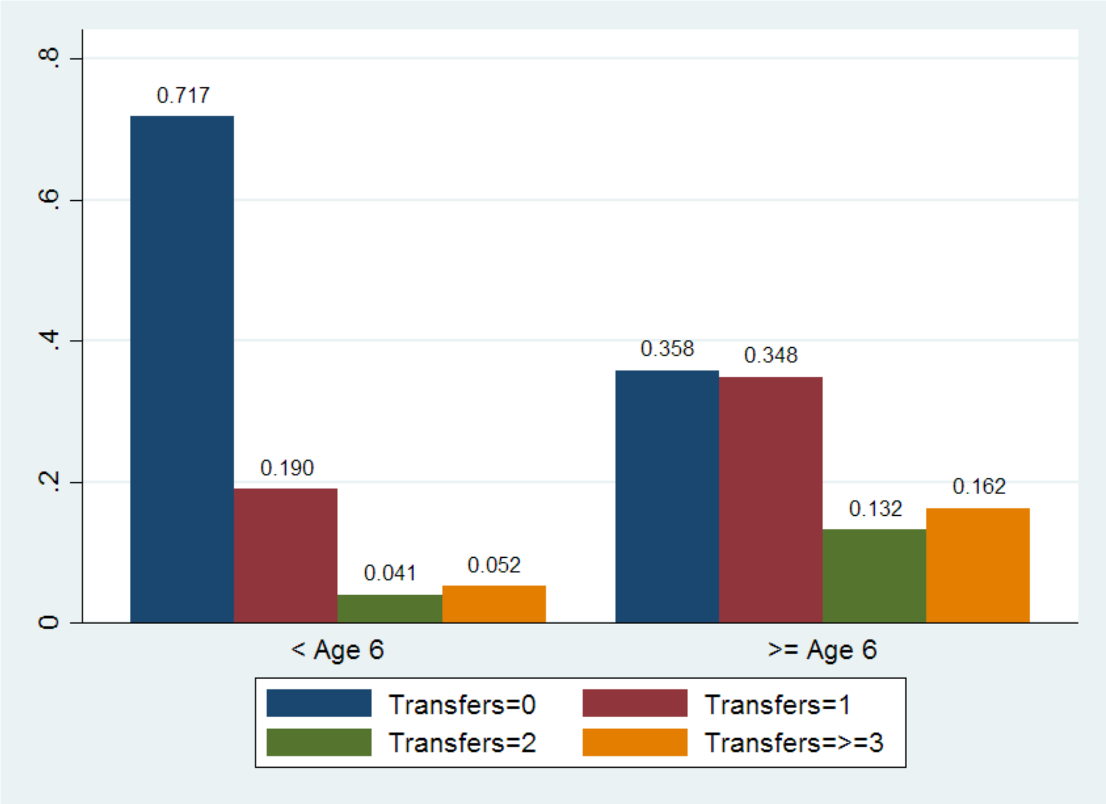

Figure 3. School transfer frequencies for migrant children based their age coming to the county/district they currently live in.

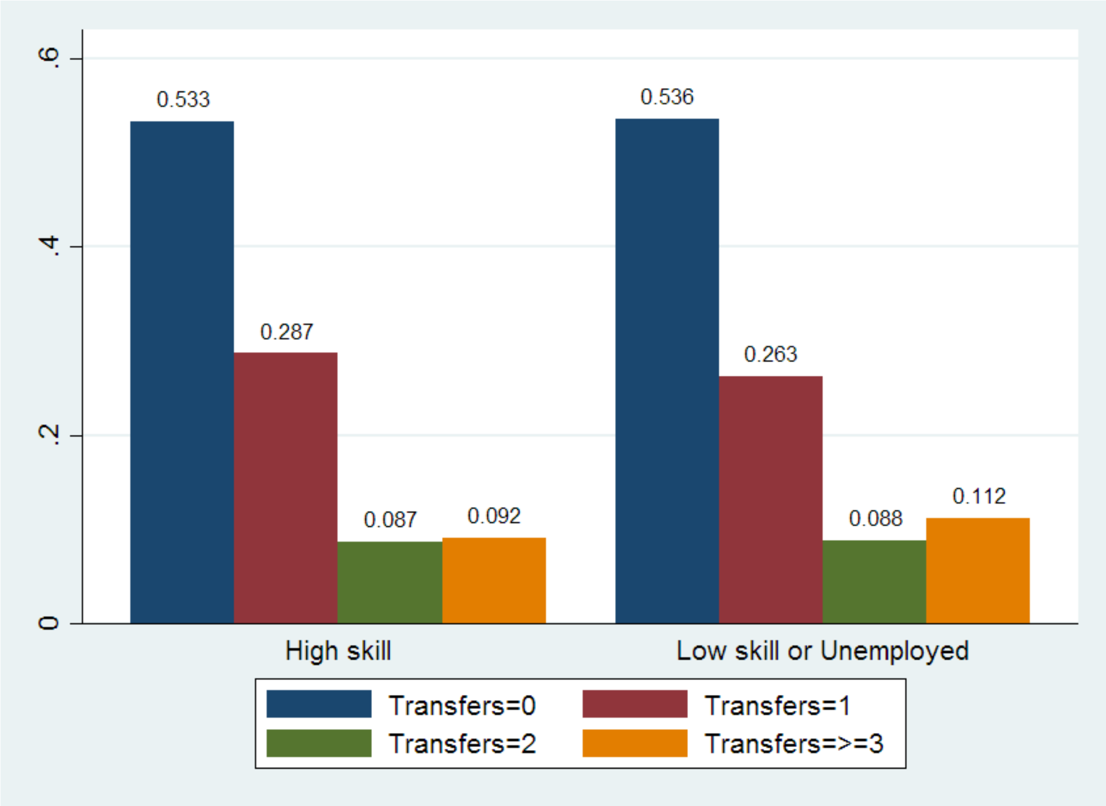

Figure 4. School transfer frequencies of migrant children based on father's occupation skill requirement.