10-9-2009

What determines basic school attainment in developing countries? Evidence from rural China

Meng Zhao
University of Minnesota, zhaox057@umn.edu

Paul Glewwe
University of Minnesota, pglewwe@umn.edu

In press from:
Meng Zhao, Paul Glewwe, What determines basic school attainment in developing countries? Evidence from rural China, Economics of Education Review, In Press, Accepted Manuscript, Available online 9 October 2009, ISSN 0272-7757
DOI: 10.1016/j.econedurev.2009.10.008
URL: http://www.sciencedirect.com/science/article/B6VB9-4XDMHNB-3/2/c813488f4ce4baf3e8964db31d116890

What determines basic school attainment in developing countries? Evidence from rural China

Abstract

This paper analyzes recent household survey data from Gansu, a less developed province in Northwest China, to examine school attainment in a poor rural area of China. Censored ordered probit regressions are used to estimate the determinants of years of schooling. Child nutritional status, as measured by height-for-age Z-scores, and household income have positive effects on completed years of schooling. Mothers' education and attitudes toward children's education also have strong effects. Children of mothers with 6 years of primary education will go to school 1.4 years longer than their counterparts whose mothers who have no education. Science labs in lower secondary schools appear to have positive impacts; providing a science lab is estimated to extend years of schooling by 1.8 years. Finally, teachers' experience in lower secondary schools also has a strong positive impact on school attainment.

Keywords

Demand for schooling, economic development, child nutrition, China

Comments

In press from:

Meng Zhao, Paul Glewwe, What determines basic school attainment in developing countries? Evidence from rural China, Economics of Education Review, In Press, Accepted Manuscript, Available online 9 October 2009, ISSN 0272-7757

DOI: 10.1016/j.econedurev.2009.10.008

URL: http://www.sciencedirect.com/science/article/B6VB9-4XDMHNB-3/2/c813488f4ce4baf3e8964db31d116890

What determines basic school attainment in developing countries? Evidence from rural China

Meng Zhao*, **Paul Glewwe**

Department of Applied Economics, University of Minnesota, St. Paul, MN 55108

Received 21 August 2007; accepted 5 February 2009

Abstract

This paper analyzes recent household survey data from Gansu, a less developed province in Northwest China, to examine school attainment in a poor rural area of China. Censored ordered probit regressions are used to estimate the determinants of years of schooling. Child nutritional status, as measured by height-for-age Z-scores, and household income have positive effects on completed years of schooling. Mothers' education and attitudes toward children's education also have strong effects. Children of mothers with 6 years of primary education will go to school 1.4 years longer than their counterparts whose mothers who have no education. Science labs in lower secondary schools appear to have positive impacts; providing a science lab is estimated to extend years of schooling by 1.8 years. Finally, teachers' experience in lower secondary schools also has a strong positive impact on school attainment. [*JEL classification*: D10; I21; O15]

Keywords: demand for schooling, economic development, child nutrition, China

1. Introduction

At the United Nations in September 2000, 189 countries agreed to the Millennium Development Goal that, by 2015, every child should complete primary education. This consensus reflects the view of most economists and international development agencies that education promotes economic growth and social development (Glewwe and Kremer, 2006).

* Corresponding author. Tel.: 651-216-1196.
E-mail addresses: zhaox057@umn.edu (M. Konishi); pglewwe@umn.edu (P. Glewwe).

China has made significant strides in promoting basic education. Its primary gross enrollment rate was 115% in 2002, compared to an average of 105% in the other East Asian developing countries. Despite this progress, China's secondary gross enrolment rate was only 70% in 2002, well below the East Asian average of 91% (World Bank, 2004). This reflects the fact that many children in China's rural areas, where 800 million Chinese reside, leave school at the end of primary education or during lower secondary school. With low skills and not even a lower secondary school diploma, they will most likely have relatively low levels of income, living conditions and social status.

There are few studies of educational attainment in China. Studies in Chinese usually use aggregate data and analyze only school enrollment. A few studies in English use either household survey or census data. They usually find that urban location, parental (particularly mother's) education, household income and the child's sex are important determinants of enrollment (Knight and Li, 1993, 1996; Hannum, 1999, 2003; Connelly and Zheng, 2003). Brown and Park (2002) found that women's empowerment and school infrastructure (e.g. rainproof classrooms) increase a child's probability of being enrolled. Yet no study has simultaneously considered the effects of parental education, child health, parental attitudes and school quality. Moreover, due to limited data omitted variable bias is a potentially serious problem in almost all previous studies.

This paper uses data from the Gansu Survey of Children and Families (GSCF) which was implemented in 2000 and 2004 in Gansu, one of China's poorest provinces, to estimate the determinants of school attainment in rural China. Gansu's per capita GDP in 2003 was 5970 Yuan (about $750), about 30% below China's average per capita GDP of 8500 Yuan (about $1067). The analysis estimates the impacts of factors such as child nutritional status, parental attitudes toward education, and school quality, on school attainment.

The GSCF is the most comprehensive data set on education in rural China, so this study is less likely to suffer from omitted variable bias. In addition, it contributes to the literature on education in China in three ways. First, studies of the impact of child health on educational outcomes in China are extremely rare. While China has made rapid progress in improving child nutrition (UNICEF, 2006), a recent study found that 40% of children in poor rural areas are stunted (Park and Zhang, 2000). This paper is the first to study the impact of nutritional status, measured by height-for-age, on child schooling in China.

Second, while the impact of parental education on children's education is often studied, studies of the impact of parental *attitudes* on children's education in developing countries are rare. The data used in this study include information on parents' preferences on their children's education, expressed separately by mothers and fathers.

Third, school quality clearly plays an important role in educational attainment because it directly affects a child's learning and, ultimately, the number of years a child attends school (see Glewwe and Kremer, 2006). Regarding China, Brown and Park (2002) found that rainproof classrooms increase years of primary education, but the probability of dropping out of primary school increases with the percentage of teachers with post-secondary education. They speculate that the unexpected negative impact of teachers' education on enrollment reflects negative correlation between teachers' education and experience, but could not test this due to lack of data on teachers' experience. This paper examines the impacts on school attainment of a comprehensive set of school-quality attributes at both the primary and secondary levels.

2. Empirical specification and variables

Households make decisions on child education by weighing the expected returns against the costs of additional years of schooling (Glewwe, 2002). A household's demand for years of schooling, denoted by *S*, is a function of the factors that influence the benefits and costs of additional schooling. There are four types of factors: child personal characteristics (***PC***), household characteristics (***HC***), community characteristics (***CC***), and school and teacher characteristics (***SC***). This can be expressed as:

$$(1) \qquad S = f(\boldsymbol{PC}, \boldsymbol{HC}, \boldsymbol{CC}, \boldsymbol{SC}) + \varepsilon,$$

where ε measures the components in ***PC***, ***HC***, ***CC***, and ***SC*** that are not in the data, as well as measurement error in *S* (implications of the nature of ε for estimation are discussed below).

The demand for education, *S*, is a continuous but unobserved variable, while observed years of schooling is determined by *S*. To estimate equation (1), this paper uses a censored ordered logit econometric specification. As in a standard (uncensored) ordered logit, the censored ordered logit uses observed data on years of schooling to estimate the unobserved demand for education, but does so in a way that lets one to retain in the sample children who are still in school. For such children, one must adjust the standard ordered logit because observed years of schooling is "right-censored"; that is, one has only a lower bound on their final years of schooling. The details of this adjustment are explained in the appendix. The intuition is that, for a child currently in school, say in grade 7, one knows that final years of schooling must be greater than 6. Most children in the Gansu data are still in school, so most observations are censored; failure to account for this would yield biased parameter estimates.

Returning to equation (1), consider each set of explanatory variables. Ideally, the vector ***PC*** should include all child characteristics that affect the demand for education, including the child's motivation, tastes for schooling, and innate ability. These variables are typically omitted because they are difficult to measure and, therefore, are combined with ε in equation (1). Some studies use observed child variables as proxies for unobserved child characteristics. For instance, innate ability may be partly explained by differences in gender, age, or nutritional status (clinical studies show that child nutrition raises cognitive skills). Improved learning ability may also increase the return to education and enhance children's motivation. We follow this approach and include sex, age, and child nutritional status (measured by height-for-age).

Many household characteristics (***HC***) can also affect child schooling. The following household variables are used in this study: education levels of the child's parents, parental attitudes toward female education and their education level for their child (reported separately by each parent), household per capita expenditures, and the amount of land farmed. Parents' education is important because it determines their ability to assist their children with schoolwork. It may also reflect their children's genetic ability. Parental educational attitudes reflect the heterogeneity in household preferences, e.g. the time discount rate or weights put on child education in the household utility function. Low income can reduce the demand for education if budget constraints limit parents' ability to invest in education.

Child education can be influenced by community factors (***CC***) , such as availability of schools, direct and opportunity costs of education, local tastes for education, and the return to education. Variables that measure school availability are the distances between the child's home and the closest primary, lower secondary and upper secondary schools. These may also capture the impacts of other community characteristics on schooling, such as the value the community places on education. The percentage of secondary degree holders among the community's adults can also be used to measure community norms regarding education. The wage for unskilled labor is used to measure the opportunity cost of education. The GSCF also has data on school fees, which are direct costs of education. A final important point is that, since the data are from 20 counties, unobserved county characteristics that affect the demand for education can be accounted for by including county fixed effects in the regression.

School and teacher characteristics (***SC***) play key roles in determining the demand for education. Better school infrastructure and more skilled teachers could increase student learning, "producing" more knowledge via a learning production function. Ultimately, better schools and teachers increase the return to investments in education.

The GSCF collected detailed information on the schools and teachers of the children in the sample, including: tuition and fees, percentage of leaking classrooms, existence of a science laboratory and/or a library, teachers' monthly salaries and annual bonuses, percentage of teachers with post-secondary education, and percentage of teachers with over 5 years of experience.[1] Brown and Park (2002) analyzed the impact on educational outcomes of several school variables. This study goes further by including information on both primary and secondary schools for all children; thus the regressions include data on the schools each child attended before (for children currently in secondary school) or the schools they are likely to attend in the future (for children in primary school), to capture more fully all relevant school characteristics

[1] In addition to salaries, teachers receive bonuses based on school principals' assessment of their performance.

3. Identification

If all the explanatory variables are accurately measured and exogenous in the sense that none are correlated with ε, a censored ordered logit provides unbiased estimates of the determinants of years of schooling. Yet omitted variables or measurement error can induce correlation between ε and the observed variables, particularly child nutritional status and household expenditures. Child nutrition, measured by height-for-age, is likely to be endogenous because it could reflect households' unobserved attitudes toward their children's quality of life. For example, parents may allocate food or medicine among children based on their characteristics, such as innate ability. If more unobserved resources are allocated to less talented children, the estimated impact of child nutrition on years of schooling will be biased downward. A two-step instrumental variable method to estimate probit and logit models, and a test for endogeneity, was introduced by Rivers and Vuong (1988). This is implemented using food prices in 2000 and children's birth weight as instruments for height-for-age.[2]

Household surveys almost always measure incomes with a large amount of error, so household expenditures (instead of income) are often used to measure households' economic resources. Yet even expenditures may be endogenous. For example, child labor could affect both household income (or expenditure) and years in school. None of the sample children worked in 2000, so household expenditure per capita in 2000 is used to measure household resources. To avoid attenuation bias due to measurement error in the expenditure variable, household durable goods (excluding items pertaining to education, such as desks and radios) are used as instruments for household expenditures.

[2] To be specific, we use village prices of rice, beans, flour and potatoes, which have the most explanatory power.

Parents' attitudes, as measured by their reported desired level of education for their children and scores on attitudes toward female education, may also be endogenous. This arises because parents may alter reported opinions regarding their children's education; for example, parents may be ashamed to report a low expectation for children who dropped out early. To check this potential problem, we compared mothers' answers in 2000 and 2004 to the exact same questions. The correlation between opinion changes and children's dropping out is very small, from 0.004 to 0.020, suggesting little to worry about for these variables.

If children or parents can influence school quality, school and teacher characteristics may reflect unobserved parents' attributes or child innate ability, leading to biased estimates. Yet children in China are usually assigned to the primary and lower secondary schools closest to their homes. There may be exceptions in urban areas, where parents are very concerned about school quality. But most rural parents have little control over school quality. In the data from Gansu, 98% of the children attended the closest primary school, and the analogous figure for lower secondary school was 95%.

4. Data and descriptive statistics

This paper uses data from the Gansu Survey of Children and Families (GSCF), which were collected in 2000 and 2004 for a random sample of two thousand children in rural areas of Gansu province who were 9 to 12 years old in the year 2000. Of these children, only 9 never enrolled in school, and of the 1991 enrolled before 2000 only 19 left school before 2000. In contrast, 225 left school between 2000 and 2004. Thus, 88% of the sample, who were age 13-17 in 2004, were still enrolled in school in that year. Figure 1 shows the percentage of children still enrolled in 2004, by age, and Table 1 shows the highest grade attained for the 225 dropouts.[3] Most dropouts left school during or immediately after the third (last) grade of lower secondary school (27.3%), followed by the fifth (second to last) grade of primary school (21.3%). Large gender differences occur at grades four and five of the primary level, with girls dropping out at double or triple the rate of boys.

[Figure 1 here]

[Table 1 here]

Table 2 shows the reasons, reported by parents, for dropping out. The reason most parents cite is that their child is unwilling to attend school (42%), followed by financial difficulties (32%) and academic difficulties (12%). The number of girls who left school due to financial difficulties is twice that of boys, while boys who dropped out of schools because of study difficulties is double that of girls. The response that the child is unwilling to attend school may not reflect the real reasons; some parents may hesitate to report financial problems or poor academic performance.

[Table 2 here]

Table 3 provides descriptive statistics for the 1918 children used in the analysis. Height-for-age Z-scores are a commonly used measure of a child's nutritional status since birth. They primarily reflect malnutrition in early childhood, so it the child's nutritional status in the first 2-3 years of life. The height for-age Z-score is defined as the number of standard deviations that a person's height is away from the median height of a reference population of healthy children of the same age and sex. A Z-score below -3 reveals severe stunting, a Z-score between -3 and -2 indicates moderate stunting, and Z-scores greater

[3] Of the 2000 children sampled in 2000, the families of 82 were not re-interviewed in 2004: 71 families moved out of the township, 8 children died, 2 children's parents divorced, and one household refused to be interviewed. No information is available for these children. The low attrition rate (4.1%) suggests little sample selection bias.

between -2 and -1 suggest mild stunting. Our data show that many children in rural areas of Gansu are malnourished; the average height-for-age Z-score is -1.2.

[Table 3 here]

5. Estimation results

This section presents estimation results, using the methodology described in Sections II and III. First, the exogeneity of the height-for-age Z-score and household expenditures was tested, using Rivers and Vuong's procedure. The null hypothesis of exogeneity was rejected for both child height-for-age Z-score (p-value = 0.018) and expenditures (p-value = 0.024).

Table 4 presents first stage estimates for the two endogenous variables, height-for-age Z-score and household expenditures. Prices of staple foods in 2000 are used as instruments for height-for-age and durable assets are used to instrument household expenditures. The instrumental variables in the first column in Table 4 exclude birth-weight, while the other two columns include it. In general, the instrumentals have the expected effects. Higher prices reduce the demand for staples and, ultimately, reduce child nutrition. Older children are more likely to be malnourished, which may reflect China's rapid economic growth; younger children benefited from higher family incomes, and perhaps better health facilities, in their first years of life. Durable assets have strong explanatory power for household expenditure but not for child nutrition. Note that the sample size falls by half due to missing price data.

[Table 4 here]

Table 5 reports censored ordered logit estimates of the determinants of years of schooling.[4] Six specifications are reported. The first is the simplest. The second adds variables that measure parental attitudes and expectations regarding education. The third and fourth repeat the first and second, respectively, but are estimated for the smaller sample with complete data on all instrument variables. The fifth presents instrumental variable (IV) estimates, using Rivers and Vuong's method, where food prices are instruments for height-for-age and household durables are instruments for household expenditure. Finally, the sixth regression adds birth-weight as an instrument for height-for-age. The results from both the fifth and sixth regressions are shown to address criticisms of using nutritional status in an early life stage (here, birth-weight) as an IV for later nutritional status (Alderman, Behrman, Lavy, Menon, 2001).

[Table 5 here]

For all regressions sown in Table 5, child age, height-for-age, mother's education, mothers' expectation of child education, per capita expenditure, distances between the closest schools and the child's home and some school quality variables all have significant explanatory power for years of schooling attained. In contrast, there is little evidence that father's education and school tuition affect children's educational attainment.

Better nourished children (measured by height-for-age) complete more years of schooling. The size of the effect is much larger in the IV regressions. The tiny differences between columns 1 and 3, and between columns 2 and 4, show that the smaller sample size does not explain differences between the results with and without instrumental variables. There are two possible reasons why the regressions without instrumental variables

[4] We also have run a censored ordered probit and a simple probit which uses the enrolment status of dropout or not as the dependent variable. The findings are similar. We report the estimates obtained in the censored ordered logit regressions because they have slightly higher significance levels than the ordered probit.

underestimate the impact of HAZ. The first could be a negative correlation between height-for-age and the error term in the underlying demand for education. Parents with relatively healthy children may spend less time helping them with their schoolwork in order to spend more time with their less healthy children. Without instrumenting HAZ, the estimated impact of HAZ is biased downward. The other possibility is measurement error in HAZ, which (if random) generally leads to attenuation bias. Measurement error is common for health and nutrition variables. For example, child height varies not only because of early childhood malnutrition but also because some children are naturally taller or shorter than others.

Based on the results in the fifth regression, a one standard deviation increase in HAZ (i.e., increasing the sample mean of the height-for-age Z-score from -1.2 to -0.2, where 0.0 is the average for a healthy population) raises school attainment by over 0.2 years.[5] When birth weight is also used as an instrumental variable in the sixth regression, the estimate for HAZ is smaller, suggesting that birth weight may violate the exclusion restriction required of instrumental variables, as some authors have claimed. Thus the fifth regression in Table 5 is our preferred specification.

Mother's education has a significant positive impact on years of schooling, and this impact is larger for girls than for boys. The impact is somewhat smaller in the IV regressions, with no difference by sex. Using the estimates from the fifth regression, an additional year of mother's education keeps her child in school 0.14 years longer. However, father's education has little impact on child schooling. Fathers probably spend less time than mothers helping children with school. The GSCF data show that fathers, on average, spent 0.68 hours helping children with homework while mothers spent 0.92 hours per week.

---

[5] The impact of HAZ on school attainment depends not only on the parameter estimates in Table 5 but also on the estimated threshold parameters, which are not shown in Table 5.

The variables measuring parental attitudes toward girls' education are from parents' responses to a series of questions. After controlling for the highest educational level mothers want for their children, the other parental attitude variables are insignificant. The impact of mothers' expectations for their children's education falls when the sample size drops (see columns (2) and (4)), but increases and is highly significant in the fifth regression, where staple food prices (but not birth weight) are used as instruments for HAZ.

Both the size and the statistical significance of the impact of per capita expenditure vary across regressions. A dramatic change from the first regression to the second suggests that the impacts of expenditures are over-estimated in the first due to omitted parental attitudes (e.g. wealthier parents may have higher tastes for child education). Yet the impacts of expenditures are higher and more significant in IV regressions, suggesting that measurement error in expenditures causes substantial attenuation bias.

The distance from the household to the closest lower secondary school reduces years of schooling in columns 1-4. The farther the distance, the sooner a child will leave school. Yet the distance to the closest upper secondary school seems to increase years in school, although significantly so only in the first two regressions. These regressions do not control for upper secondary school characteristics, so this distance may reflect some unobserved aspects of school or teacher quality. In any case, the preferred IV specification (column 5) shows no effect for all school distance variables; presumably more accurate estimation of the impact of household expenditures removes bias in the estimated impacts of these variables.

The impacts of both primary and lower secondary tuition are statistically insignificant. As a further check, an interaction term, the product of household expenditure and tuition, was added to see whether the impacts of tuition vary by household wealth (not shown in Table 5), but this interaction was statistically insignificant. There are several reasons why tuition may have little effect. The estimated impacts may be biased because tuition could be positively correlated with unobserved school quality, which would underestimate the impact of tuition. Another concern is measurement error in reported tuition, which may lead to downward bias in the estimation. We also jointly test all the primary school characteristics and reject the null hypothesis that none has any impact on the years of schooling (p-value = 0.016).

One potentially important finding in Table 5 is that lower secondary school science labs appear to have a positive, statistically significant impact on school attainment. Many students aged 13-17 start to develop their life and career plans. Some may lose interest in school if teaching is not interactive. Science labs can help students relate abstract textbook theory to reality and thus increase students' interest in science. Science labs may also help students learn by trial and error, instead of by rote learning, which may increase students' learning and thus reduce dropping out due to poor school performance.

Yet this interpretation of the estimated effect of science labs can be disputed; science labs may be correlated with unobserved school characteristics. To check further, Table 6 compares basic school and community characteristics for lower secondary schools with and without science labs. While schools with science labs tend to be larger, with more resources, the strong impact of science labs on years of schooling remains when all variables in Table 6 are added as regressors. The estimated effect of science labs is very large; a child in a school with a science lab completes 1.8 more years of school (using estimates in Table 5, column 5). Table 7 presents a simple cost-benefit calculation of building a science lab in a lower secondary school. It shows a huge 8.8 million yuan benefit at a cost of only 40,000 yuan.

[Table 6 here]

[Table 7 here]

Finally, the percentage of teachers with 5 or more years of teaching experience in lower secondary school also has a significant positive impact on years of schooling. This is consistent with Brown and Park's (2002) conjecture; they could not provide evidence due to lack of data on teacher experience. However, the impact of teacher experience in primary school is negative (though significant only for the IV results). A possible explanation is that younger teachers, who typically have fewer years of experience, may be more energetic, innovative and willing to try new ideas, so which helps them better motivate primary school students. The higher experience of older teachers may also increase their effectiveness, but our results suggest that this is outweighed by the greater enthusiasm of younger teachers.

6. Conclusion

This paper has attempted to identify the determinants of educational attainment in rural areas of Gansu province, one of China's poorest provinces. An unusually rich set of household and school data allowed us to conduct a comprehensive analysis that accounts for many factors that have been ignored in previous research. The data confirm that mother's education and household income have strong positive impacts on years of schooling. In addition, mothers' expectations of the highest education level for their children increase their years of schooling. Children's nutritional status also has a strong positive effect. Finally, science labs and teacher experience in lower secondary schools appear to increase students' years in school.

These findings suggest that policies that target child nutrition, and perhaps school science labs, should raise school attainment and so reduce dropping out in rural China. There will also be added positive effects for future generations via the impact of maternal education on child schooling. Yet much more remains to be learned. First, research is needed on how best to improve child nutrition in rural China. Second, investigation of the factors that shape mothers' expectations for their children would be useful. Third, the role of science labs should be checked further to determine whether they really make school more attractive; the most convincing way to do this is to conduct a randomized trial that provides science labs to randomly selected schools. Finally, more must be learned about how to make schools more effective in promoting children's learning of the curriculum; once children are in school they must learn the skills they will need to improve their lives after they finish their schooling.

Acknowledgments

We would like to thank an anonymous referee for comments on an earlier draft of our paper.

References

Alderman, H., Behrman, J.R., Lavy, V., & Menon, R. (2001). Child health and school enrollment: a longitudinal analysis. *Journal of Human Resources, 36,* 185-205.

Brown, P., & Albert P. (2002). Education and poverty in rural China. *Economics of Education Review, 21(6),* 523-541.

Connelly, R., & Zheng, Z. (2003). Determinants of school enrollment and completion of 10 to 18 year olds in China. *Economics of Education Review, 22(4),* 379-88.

Glewwe, P., & Jacoby, H. (1994). Student achievement and schooling choice in low income countries: evidence from Ghana. *Journal of Human Resources 29(3),* 843-864.

Glewwe, P. (2002). Schools and skills in developing countries: education policies and socioeconomic outcomes. *Journal of Economic Literature 40(2),* 436-482.

Glewwe, P., & Kremer, M. (2006). Schools, teachers and education outcomes in developing countries. In E. Hanushek & F. Welch, eds. *Handbook of the Economics of Education*. North Holland.

Hannum, E. (1999). Political change and the urban-rural gap in basic education in China, 1949-1990. *Comparative Education Review 43(2),* 193-211.

Hannum, E. (2003). Poverty and basic education in rural China: villages, households, and girls' and boys' enrollment. *Comparative Education Review 47(2).*

Knight, J., & Li, S. (1993). The determinants of educational attainment in China. In K. Griffin, & R. Zhao, Eds. *The Distribution of Income in China* (pp. 285-330). London: Macmillan Press.

Knight, J., and Li, S. (1996). Educational attainment and the rural-urban divide in China. *Oxford Bulletin of Economics and Statistics 58(1),* 83-117.

Maddala, G.S. (1983). *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

Park, A., & Zhang, L.X. (2000). Mother's education and child health in China's poor areas. Mimeo, University of Michigan Department of Economics.

Rivers, D., & Vuong, Q.H. (1988). Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of econometrics 39*, 347-366.

UNICEF. (2006). Progress for children: a report on nutrition. UNICEF: New York, NY.

World Bank. (2004). *World Development Report 2004: Making Services Work for Poor People.* Washington, DC.

Appendix: the censored order probit and logit models

Following Maddala (1983) and Glewwe and Jacoby (1994), assume that *f* in equation (1) is a linear function of a vector **x** that includes personal characteristics ***PC***, household characteristics ***HC***, community characteristics ***CC***, school and teacher characteristics ***SC***. For each person *i* (*i = 1,..., N*), the demand for education is a linear function of x, $S_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i$, where $\boldsymbol{\beta}$ is a vector of coefficients associated with all the variables in **x**. The observed years of schooling, y_i, is assumed to follow:

$$y_i = 0 \text{ if } -\infty \le S_i < \alpha_0 \quad \Leftrightarrow \quad -\infty \le \mathbf{x}_i'\boldsymbol{\beta} + u_i < \alpha_0 \quad \Leftrightarrow \quad \varepsilon_i < \alpha_0 - \mathbf{x}_i'\boldsymbol{\beta}$$

$$y_i = 1 \text{ if } \alpha_1 \le S_i < \alpha_2 \quad \Leftrightarrow \quad \alpha_1 \le \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i < \alpha_2 \quad \Leftrightarrow \quad \alpha_1 - \mathbf{x}_i'\boldsymbol{\beta} < \varepsilon_i < \alpha_2 - \mathbf{x}_i'\boldsymbol{\beta}$$

……………

$$y_i = m-1 \text{ if } \alpha_{m-1} \le S_i < \alpha_m \quad \Leftrightarrow \quad \alpha_{m-1} \le \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i < \alpha_m \quad \Leftrightarrow \quad \alpha_{m-1} - \mathbf{x}_i'\boldsymbol{\beta} < \varepsilon_i < \alpha_m - \mathbf{x}_i'\boldsymbol{\beta}$$

$$y_i = m \text{ if } \alpha_m \le S_i < \infty \quad \Leftrightarrow \quad \alpha_m \le \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i \quad \Leftrightarrow \quad \alpha_m - \mathbf{x}_i'\boldsymbol{\beta} < \varepsilon_i$$

with m the highest level of y_i and α's the underlying cut-offs that determine, jointly with S_i, the observed years of schooling. Assume that ε_i is independent and identically distributed (i.i.d.) for all i; the probability of observing y_i at each level can be expressed as the follows:

$$\Pr(y_i = 0 \mid \mathbf{x}_i) = F(\alpha_0 - \mathbf{x}_i^{'}\boldsymbol{\beta})$$

$$\Pr(y_i = 1 \mid \mathbf{x}_i) = F(\alpha_1 - \mathbf{x}_i^{'}\boldsymbol{\beta}) - F(\alpha_0 - \mathbf{x}_i^{'}\boldsymbol{\beta})$$

………………

$$\Pr(y_i = m-1 \mid \mathbf{x}_i) = F(\alpha_m - \mathbf{x}_i^{'}\boldsymbol{\beta}) - F(\alpha_{m-1} - \mathbf{x}_i^{'}\boldsymbol{\beta})$$

$$\Pr(y_i = m \mid \mathbf{x}_i) = 1 - F(\alpha_m - \mathbf{x}_i^{'}\boldsymbol{\beta})$$

where F is the c.d.f. of ε_i. If ε_i follows a normal (logistic) distribution, this is an ordered probit (logit). If $y_i = j$ ($j = 0,\ldots,m$) and is censored (e.g. person *i* is currently enrolled in level *j*), all we know is that her final years of schooling will be greater than or equal to (j - 1). Thus the probability of observing j years of schooling should be $\Pr(y_i = j \mid \mathbf{x}_i) = 1 - F(\alpha_j - \mathbf{x}_i^{'}\boldsymbol{\beta})$. Letting $I_{ij} = 1$ if $y_i = j$, $I_{ij} = 0$ otherwise; $d_i = 1$ if y_i is censored and $d_i = 0$ otherwise, the log likelihood of observing the whole sample of size N can be expressed as:

$$\ln L(\boldsymbol{\alpha},\boldsymbol{\beta}) = \Sigma_{i=1}^{n} \ln \Pr(y_i \mid \mathbf{x}_i).$$

$$\ln L(\boldsymbol{\alpha},\boldsymbol{\beta}) = \Sigma_{i=1}^{n}\Sigma_{j=1}^{m} I_{ij}\{\ln[F(\alpha_j - \mathbf{x}_i^{'}\boldsymbol{\beta})]^{1-d_i} - F(\alpha_{j-1} - \mathbf{x}_i^{'}\boldsymbol{\beta})\}.$$

Censored ordered probit (logit) of estimators for the **α**'s and **β**'s are those that maximize the above log likelihood function, assuming that ε_i follows a normal (logistic) distribution.

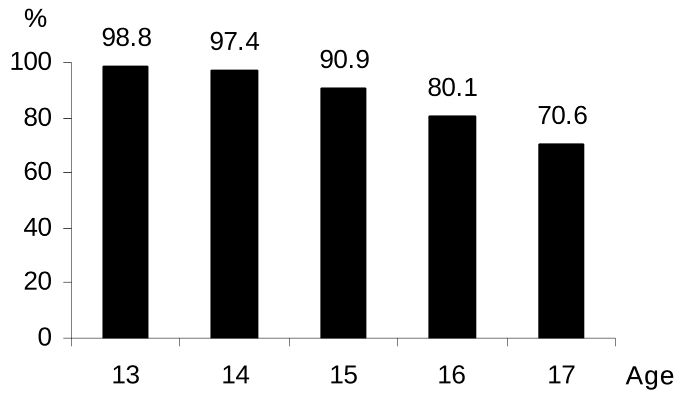

Figure 1: Enrollment rate by age

Table 1
Highest grades attained by dropouts

	Grade 1, primary	Grade 2, primary	Grade 3, primary	Grade 4, primary	Grade 5, primary	Grade 6, primary	Grade 1, lower secondary	Grade 2, lower secondary	Grade 3, lower secondary	Grade 1, upper secondary	Total
Boys	1	-	3	7	12	5	20	20	30	2	100
Girls	1	-	3	12	36	8	21	14	30	0	125
Total	2	-	6	19	48	13	41	34	60	2	225

Table 2
Reasons for dropping out, as reported by households

	Financial difficulty	Academic difficulty	Child does not want to go to school	Finished the desired	Other reasons	Total
Boys	23	17	41	9	12	102
Girls	52	9	56	4	11	132
Total	75	26	97	13	23	234

Table 3
Description of variables used

Variable	Mean	S.D.	Min	Max
Dependent variable				
Grade of dropping out	6.71	2.1	0	10
Current grade for those who haven't dropped out	7.25	1.8	1	15
Personal characteristics				
Sex (=1 if boy)	0.57	0.5	0	1
Age	15	1.2	13	20
Height-for-age z-score	-1.2	1.1	-6.8	1.7
Family characteristics				
Years of fathers' education	7	3.6	0	15
Years of mothers' education	4.3	3.5	0	13
Father's attitude towards female edu.	1.3	1.0	-3	3
Mother's attitude towards female edu.	3.1	1.2	-3	4
Highest grade father wishes child to achieve*[1]	3.6	0.8	1	4
Highest grade mother wishes child to achieve	3.8	0.6	1	4
Amount of land (mu*[2])	9.6	7.2	0	80
Family size	4.7	1.1	2	10
Per capita expenditure in 2000 (yuan/yr)	1418	391	130	13,875
Community characteristics				
Distance to primary school (km)	0.9	2.2	0	12
Distance to lower secondary school (km)	3.7	4.2	0	30
Distance to upper secondary school (km)	12	12.8	0.3	80
Primary school/teacher characteristic				
Tuition and fees (yuan/semester)	127	107	0	2017
Leaking classrooms (%)	0.2	0.3	0	1
School has science lab (=1 if yes)	0.25	0.4	0	1
School has library (=1 if yes)	0.69	0.5	0	1
Teachers with 5+ years' experience (%)	0.79	0.2	0	1
Teachers with post-secondary degree (%)	0.84	0.2	0	1
Teacher bonus (yuan/yr)	66	103	0	1,600
Teacher salary (yuan/month)	573	155	0	1,040
Lower secondary school/teacher characteristics				
Tuition and fees (yuan/semester)	404	522	19	8,200
Leaking classrooms (%)	0.19	0.3	0	1
School has science lab (=1 if yes)	0.52	0.5	0	1
School has library (=1 if yes)	0.93	0.3	0	1
Teachers with 5+ years' experience (%)	0.81	0.1	0.2	1
Teacher bonus (yuan/yr)	176	249	0	1,800
Teacher salary (yuan/month)	957	181	0	1,825

*[1] Code for education level: 1-Primary; 2-Lower secondary; 3-Upper secondary; 4-Vocational school; 5-Post secondary training; 6-Two-year college; 7-Four-year college; 8-Graduate.
*[2] 1 mu = 667 square meters.

Table 4
Least square regressions explaining HAZ and expenditures

	HAZ	HAZ	Expenditures
Instrumental variables			
Price of rice in 2000	-1.160 *	-1.102 *	-0.523
	(0.632)	(0.643)	(0.237)
Price of beans in 2000	0.221	0.169	0.045
	(0.204)	(0.206)	(0.073)
Price of flour in 2000	-0.574 **	-0.618 **	-0.019
	(0.231)	(0.241)	(0.085)
Price of potatoes in 2000	-0.259 ***	-0.285 ***	0.122 ***
	(0.098)	(0.101)	(0.037)
Child birth-weight	-	0.087 **	0.006
		(0.410)	(0.015)
Log of durable goods	0.026	0.022	0.228 ***
	(0.048)	(0.049)	(0.018)
Personal characteristics			
Sex (1=boy)	0.112	0.088	0.109
	(0.259)	(0.267)	(0.097)
Age	-0.087 **	-0.091 **	0.026 *
	(0.037)	(0.038)	(0.014)
Family characteristics			
Years of fathers' education	0.015	0.013	-0.006
	(0.018)	(0.019)	(0.007)
Years of mothers' education	0.011	0.014	0.009
	(0.019)	(0.020)	(0.007)
Interaction of child sex and father education	0.018	0.017	-0.003
	(0.025)	(0.026)	(0.009)
Interaction of child sex and mother education	-0.054 **	-0.052 *	-0.014
	(0.026)	(0.026)	(0.010)
Mothers' expectation of highest grade child to achieve	-0.060	-0.045	0.037 *
	(0.055)	(0.061)	(0.022)
Fathers' expectation of highest grade child to achieve	0.045	0.042	0.041
	(0.079)	(0.082)	(0.029)
Community characteristics			
Distance to primary school	-0.054	-0.051	0.014
	(0.037)	(0.037)	(0.013)
Distance to lower secondary school	-0.015	-0.012	-0.011 *
	(0.019)	(0.020)	(0.006)
Distance to upper secondary school	0.007	0.007	0.006 *
	(0.009)	(0.009)	(0.003)
Primary school/teacher characteristics			
Log of tuition	-0.059	-0.100	0.161 **
	(0.189)	(0.198)	(0.069)
Leaking classroom (%)	-0.052	-0.095	-0.081
	(0.155)	(0.158)	(0.058)
Science lab (1=yes)	-0.040	-0.034	-0.005
	(0.115)	(0.118)	(0.043)
Library (1=yes)	0.037	0.026	0.092 *
	(0.139)	(0.144)	(0.053)
Teacher with 5+ years of experience (%)	0.447	0.46	-0.085
	(0.343)	(0.353)	(0.121)

	HAZ	HAZ	Expenditures
Teacher with post-secondary degree (%)	0.363	0.469	0.223 **
	(0.305)	(0.311)	(0.112)
Log of teacher bonus	-0.019	-0.019	0.011
	(0.024)	(0.024)	(0.009)
Log of teacher salary	0.073 *	0.071 *	0.031 **
	(0.043)	(0.043)	(0.015)
Lower Secondary school/teacher characteristics			
Log of tuition	0.473 **	0.499 **	0.401 ***
	(0.195)	(0.199)	(0.067)
Leaking classroom (%)	-0.366 ***	-0.202	-0.01
	(0.138)	(0.152)	(0.055)
Science lab (1=yes)	0.183 *	0.163	-0.036
	(0.097)	(0.099)	(0.036)
Library (1=yes)	0.456	0.421	0.161
	(0.282)	(0.299)	(0.105)
Teacher with 5+ years of experience (%)	0.122	0	0.355 **
	(0.405)	(0.422)	(0.142)
Log of teacher bonus	0.029	0.03	0.019 ***
	(0.019)	(0.020)	(0.007)
Log of teacher salary	-0.091	-0.131	0.012
	(0.073)	(0.089)	(0.030)
Observations	711	671	745
R-squared	0.18	0.19	0.39

Table 5
Censored ordered logit regression results

	Censored ordered logit without IV				Censored ordered logit with IV	
Personal characteristics						
sex (1=boy)	0.248	0.649	0.048	-0.122	-0.140	-0.135
	(0.260)	(0.482)	(0.447)	(0.699)	(0.661)	(0.749)
age	1.375 ***	1.406 ***	1.302 ***	1.343 ***	1.298 ***	1.306 ***
	(0.100)	(0.105)	(0.147)	(0.169)	(0.147)	(0.162)
height for age z-score	0.275 ***	0.244 ***	0.243 ***	0.231 **	2.144 ***	1.671 **
	(0.060)	(0.059)	(0.090)	(0.099)	(0.373)	(0.800)
Family characteristics						
years of fathers' education	0.018	0.020	-0.002	0.001	-0.008	-0.014
	(0.029)	(0.031)	(0.039)	(0.042)	(0.034)	(0.035)
years of mothers' education	0.121 ***	0.117 ***	0.145 ***	0.138 ***	0.122 ***	0.138 ***
	(0.026)	(0.033)	(0.044)	(0.050)	(0.041)	(0.046)
Interaction of child sex and father education	0.049	0.030	0.065	0.035	0.066 *	0.056
	(0.032)	(0.037)	(0.049)	(0.055)	(0.036)	(0.049)
Interaction of child sex and mother education	-0.101 ***	-0.096 ***	-0.092 **	-0.099 **	-0.098 ***	-0.111 ***
	(0.030)	(0.034)	(0.044)	(0.049)	(0.036)	(0.040)
Father attitude towards female education	-	0.196	-	0.006	-0.042	-0.050
		(0.120)		(0.162)	(0.135)	(0.144)

	Censored ordered logit without IV				Censored ordered logit with IV	
Mother attitude towards female	-	0.030	-	0.016	0.044	0.014
education	-	(0.065)	-	(0.081)	(0.095)	(0.091)
Fathers' expectation of highest	-	0.181	-	0.153	0.119	0.023
grade child to achieve	-	(0.129)	-	(0.164)	(0.164)	(0.171)
Mothers' expectation of highest	-	0.307 ***	-	0.148	0.167 *	0.107
grade child to achieve	-	(0.075)	-	(0.128)	(0.098)	(0.125)
Log of per capita expenditure	0.341 *	0.221	0.208 **	-0.008	1.108 ***	1.045 **
in 2000	(0.174)	(0.167)	(0.105)	(0.092)	(0.243)	(0.288)
distance to primary school (km)	-0.094	-0.074	0.029	0.082	-0.008	0.012
	(0.057)	(0.059)	(0.140)	(0.128)	(0.108)	(0.132)
distance to lower secondary	-0.107 ***	-0.112 ***	-0.107 **	-0.104 **	-0.037	-0.076 *
school (km)	(0.037)	(0.043)	(0.049)	(0.046)	(0.026)	(0.043)
distance to upper secondary	0.029 ***	0.035 ***	0.026	0.024	0.016	0.018
school (km)	(0.009)	(0.011)	(0.028)	(0.025)	(0.026)	(0.026)
Primary school/teacher characteristics						
Log of tuition	-0.276	-0.336	1.635	1.444	0.523	1.251
	(0.510)	(0.579)	(1.338)	(1.087)	(1.185)	(0.926)
Leaking classroom (%)	-0.239	-0.511	-0.327	-0.648	-0.350	-0.489
	(0.319)	(0.414)	(0.506)	(0.482)	(0.538)	(0.512)
Science lab (1=yes)	0.146	0.146	0.049	0.023	-0.065	-0.09
	(0.157)	(0.169)	(0.280)	(0.255)	(0.309)	(0.228)
Library (1=yes)	0.415	0.319	0.692 **	0.792 ***	0.850 **	0.629 ***
	(0.311)	(0.299)	(0.351)	(0.290)	(0.363)	(0.241)
Teacher with 5+ years of	-0.878	-0.844	-2.329 **	-2.007 **	-1.158	-1.091
experience (%)	(0.564)	(0.589)	(0.977)	(0.932)	(0.775)	(0.849)
Teacher with post-secondary	-0.063	0.133	0.217	0.484	0.406	0.556
degree (%)	(0.526)	(0.571)	(0.567)	(0.542)	(0.549)	(0.590)
Log of teacher bonus	0.049	0.059	-0.036	-0.026	-0.003	-0.053
	(0.038)	(0.039)	(0.057)	(0.044)	(0.049)	(0.060)
Log of teacher salary	-0.135 **	-0.098 *	-0.054	-0.038	-0.096	-0.105
	(0.058)	(0.051)	(0.060)	(0.058)	(0.068)	(0.066)
Lower secondary school/teacher characteristics						
Log of tuition	0.548	0.684	0.282	0.663	0.543	0.708
	(0.423)	(0.485)	(0.474)	(0.433)	(0.466)	(0.507)
Leaking classroom (%)	-0.383	-0.357	-0.867 **	-0.685	-0.626 **	-0.541
	(0.526)	(0.562)	(0.380)	(0.452)	(0.291)	(0.350)
Science lab (1=yes)	1.622 ***	1.534 ***	1.638 ***	1.649 ***	1.507 ***	1.550 ***
	(0.257)	(0.271)	(0.451)	(0.477)	(0.387)	(0.399)
Library (1=yes)	-0.818 **	-0.864	0.447	0.678	-0.016	0.222
	(0.362)	(0.525)	(0.749)	(0.674)	(0.909)	(0.668)
Teacher with 5+ years of	1.513 *	1.211	6.483 ***	6.767 ***	5.256 ***	5.223 ***
experience (%)	(0.890)	(0.889)	(1.800)	(1.829)	(1.532)	(1.865)
Log of teacher bonus	0.032	0.009	0.014	0.003	-0.007	0.002
	(0.047)	(0.048)	(0.068)	(0.071)	(0.060)	(0.064)
Log of teacher salary	-0.003	-0.051	-0.119	-0.230	-0.151 **	-0.245
	(0.132)	(0.100)	(0.083)	(0.231)	(0.063)	(0.215)
Observations	**1487**	**1344**	**711**	**671**	**787**	**745**

Table 6
Comparisons of lower secondary schools with and without science labs

	No science lab (124 schools)		Has science labs (94 schools)	
	Mean	S.D.	Mean	S.D.
School / teacher characteristics				
Expenditure per pupil (yuan)	46	41	116 *	242 *
Whether has a library (1 = yes)	0.8	0.4	1.0	0.0
Teacher monthly salary (yuan)	900	124	999	179
Teacher annual bonus (yuan)	170	224	176	262
Tuition and textbook costs (yuan)	219	59	262	101
Total number of teachers	44	27	63	41
% of formal teachers	90	10	90	10
Community characteristics				
Average wage level in the village (yuan)	17	6	19	7
% of people with post-secondary education	10	10	20	10
% of people who are illiterate	20	20	10	10
Average (log) income per capita	7.0	0.9	7.3	0.9
% of villages that have enterprises	20	40	30	40

*The high mean and standard deviation for the expenditure per pupil in schools with science labs is driven by 6 outliers whose expenditures per pupil are over 500 yuan. Thus, comparing the medians might be more preferable: median of expenditure per pupil in schools without science lab = 45; median of expenditure per pupil in schools with science lab = 68.

Table 7
A simple cost-benefit analysis of having a science lab

Cost (Yuan)

¥40,000 to build one science lab (covering equipments, specimens, materials, etc.)

Benefit (Yuan)

(a) Income gain per pupil
Step 1: Having science lab increase years of schooling by 1.8 years;
Step 2: 1.8 years of schooling raises the income in the first working period by 1.8x104 = ¥187.2;
Step 3: Total discounted income gain in 30 years = 187.2x(1-0.95^30)/(1-0.95) = ¥2940.4
(b) Total number of pupils who benefit from a science lab
1 year: 300 graduates;
10 years: 300x10=3,000 graduates
(c) Total income gain from a science lab
= income gain per pupil x total number of pupils who benefit
= (a) x (b) = ¥2940.4 x 3,000 = ¥8.8 million

Note: The calculation is based on several rough assumptions:
1. The income gain per pupil per year of schooling is estimated in a wage equation for adults aged 20-50 who earn a regular wage in the sample. Increasing one year of schooling increase annual income by $13 (p-value = 0.06). 2. The future income discount rate = 0.05. 3. 300 students graduate per year 4. A science lab lasts 10 years.