Lancaster University Management School
Working Paper
2008/018

EARLY YEARS EDUCATION AND SUBSEQUENT SCHOOLING IN RURAL INDIA: AN ECONOMIC ANALYSIS

Geraint Johnes

The Department of Economics
Lancaster University Management School
Lancaster LA1 4YX
UK

EARLY YEARS EDUCATION AND SUBSEQUENT SCHOOLING IN RURAL INDIA: AN ECONOMIC ANALYSIS

Geraint Johnes

Lancaster University Management School
Lancaster LA1 4YX
United Kingdom

T: +44 1524 594215
F: +44 1524 594244
E: G.Johnes@lancaster.ac.uk

November 2008

ABSTRACT

This study uses LSMS microdata to evaluate the impact of early years education on subsequent educational participation in the states of Uttar Pradesh and Bihar in Northern India. It is established that, alongside a number of economic and demographic variables, pre-school education has a significantly positive impact on subsequent experience. The result is robust to correction for endogeneity bias and clustering of observations.

Keywords: education, development
JEL Classification: I20, O10

1. Introduction

In developed economies there is an increasing body of evidence which supports the idea that educational interventions made at a very young age are highly effective, and certainly more so than later remedial interventions (Heckman and Masterov, 2007). This literature has immediate relevance for countries such as India, where limited public resources for education need to be spent in such a way as to maximise the likely social return. Yet we know little about the impact of programmes that exist specifically to promote pre-school learning.

This gap in our knowledge is partly due to the availability of data. To evaluate fully the impact of early years interventions, we need longitudinal data for individuals which is collected over a period of many years, following pre-schoolers through into the labour market. Such data sets require both a substantial long term commitment of resource and a communications infrastructure that allows repeated contact with respondents. The initial investment in this type of data set comes long before the full benefits can be realised. While some developed countries have collected this type of data set, they are not particularly common. Unfortunately data of this kind are not available for India.

We do, however, have access to some cross-section data that include limited information about the educational history of individuals. This allows us to attempt an evaluation of the impact of early years education on subsequent educational participation. While this is not tantamount to an assessment of the labour market benefits of pre-school, it represents an advance on current understanding in the specific context of India. It may also serve to reinforce more general findings about the power of early years education by demonstrating that the benefits of such education are not confined to a small cluster of highly developed economies.

The remainder of the paper is structured as follows. In the next section, we survey the relevant literature. This comes in two main branches: first, the recent international work on the impact of early years education, and the second is the literature on education in India. In section 3, we describe the data source used for the analysis in this paper, and in section 4 we report on that analysis. Section 5 draws together some conclusions, along with some suggestions about the direction in which future research should go.

2. Literature review

There has been considerable investment in education in India over recent years as the country strives to attain universal primary education. This investment has included spending on centres that support very young children and their mothers – in terms of education, health and nutrition. But the role of such centres in the specific sphere of education has not been the subject of serious evaluation in the same way as has been the case in the developed countries. We begin, therefore, with a more general review of the literature on early years education.

2.1 Early years education

Much recent work in this area has been facilitated by the Early Childhood Research Collaborative (http://www.earlychildhoodrc.org/) which is in turn sponsored by the Federal Reserve Bank of Minneapolis and the Centra for Early Education and Development at the University of Minnesota. This Collaborative has organised conferences on the topic, with key papers being delivered by economists such as Jim Heckman and Flavio Cunha, Jeremy Finn and Charles Achilles, and Clive Belfield and Hank Levin. Through the involvement of the Bank and its extensive series of publications, the issue of early years education has received a great deal of publicity in the United States.

The work of Heckman has been particularly influential (Cunha *et al*., 2006; Heckman, 2008). Many of his earlier studies concern the effectiveness of a variety of remedial training programmes for low skilled workers, and the conclusions reached from these suggests that such schemes are of limited net benefit (see, for example, Heckman *et al*., 1999). More recent work, however, suggests that the economic return to investment to early years education is substantial (see, for example, Heckman and Masterov, 2007;
Cunha and Heckman, 2008). This is in spite of the fact that the delay between the time of investment and the onset of returns is long.

A persuasive explanation of why this should be so is provided by Cunha and Heckman (2007, 2008). They separate skills into cognitive and noncognitive types. The former concern knowledge acquired through learning, reasoning and experience, while the latter concern personality traits including motivation, self-discipline, diligence and attention to detail. Both cognitive and noncognitive skills are known to affect labour market earnings (Heckman and Rubinstein, 2001). Rather than estimate a production function for test scores, however, Cunha and Heckman develop a dynamic factor model capable of evaluating the impact on subsequent labour market earnings (unlike test scores a metric free measure) of investments and skill acquisition of different types at different periods of time. They account for the possibility that past acquistition of skills (of either type) affects the current effectiveness of investment in skills (of either type). They estimate their model using the Children of the National Longitudinal Survey of Youth (CNLSY/79) dataset and find, in common with many studies surveyed in Cunha *et al*. (2006), that the most productive period for parental investment in their children's cognitive skills is during the child's early years, up to the age of 8 or 9 years. For noncognitive skills, the most productive period is a couple of years later. The decline in the return to investments in cognitive skills as children age is consonant with the notion that early years learning will result in high returns.

The work referred to above has started to have considerable impact in policy circles, but has not come out of a vacuum. Earlier work by Currie and Thomas (1995) and Currie (2001) had already pointed to the effectiveness of programmes such as Head Start, aimed at disadvantaged young children in the United States. More recently, several studies have addressed the policy making community directly, either within specific local jurisdictions or at national level (see, for example, Karoly and Bigelow, 2005, and Lynch, 2007). But it is fair to say that the bulk of work in this area has been

concentrated in highly developed countries, and especially in the United States, and it is not clear to what extent the results can be generalised.

2.2. Education in India

Much water has flowed under the bridge since Stephen Hyneman (1980) challenged the orthodox view that increasing investment in education in India would be uneconomic. This view had gained currency because of high levels of unemployment amongst highly educated workers. Hyneman attributes this to overlong periods of search unemployment amongst young, educated workers. At lower levels of education, the demand for schooling remained high, suggesting that there remained substantial returns to educational investments at at least these levels. This is confirmed by Duraisamy (2002) who estimates rates of return to education in India over the period 1983-94. He finds that rates of return rise up to secondary education level, and, while they are somewhat lower for higher education, they remain well above the external rate of return. Hence, for example, the annual rate of return to college or university education in the early 1990s, based on OLS estimation of a Mincerian earnings function, was 11.7 per cent. Returns to middle and secondary education are higher for women than men, but this pattern is reversed in the case of higher education.

Despite the high rates of return, a substantial proportion of Indian children still lack access to effective basic education. Youth literacy rates, while higher than those in neighbouring Bangladesh and Pakistan, are far below the (near 100 per cent) levels achieved in Sri Lanka and China (a major comparator economy). However access to primary education has improved substantially in recent years, so that by 2007 some 95.8 per cent of 6 to 14 year olds were in school.[1] That said, access to secondary education remains poor in comparison with other countries. Facilities remain poor by international standards, although the Sarva Shiksha Abhiyan (Campaign for Education for All) and its predecessor campaign have improved matters over the last 15 years or so (Kingdon, 2007). Teacher absence has historically been high, though there are signs that this too is improving – the average attendance rate for teachers rose from 78.8 per cent to 87.4 per cent between 2004 and 2007. Over the same period there was also a slight increase in pupil attendance rates, to 75.5 per cent.[2]

State governments in India assume primary responsibility for education, and in consequence there remains a great deal of regional variation in the nature of provision (Duraisamy, 2002; Kingdon, 2007). But the typical pattern is one in which 5 years of primary education is followed by 3 years of middle schooling, 2 years of lower secondary and 2 years of upper secondary education. In spite of the government's key role, the private sector has been booming, and accounts for a substantial proportion of education at all levels. In India as a whole, some 19.3 per cent of pupils attended private schools in 2007.

[1] http://pratham.org/aser07/ASER07.pdf. This source also provides information on the quality of education, including student achievement on simple tests. But in the absence of data that are internationally comparable – India does not, for instance, participate in studies such as TIMSS or PISA – it is difficult to interpret these statistics.
[2] http://pratham.org/aser07/ASER07.pdf.

At the pre-school level, Anganwadi centres were set up during the 1970s as part of the Integrated Child Development Services (ICDS) programme. They are a means of providing care and support to disadvantaged families, especially those with young children. They provide basic health facilities, sanitation, food, and learning facilities for pre-school children. Across the whole of India, Anganwadi centres provide support to over 30 million young children and their mothers. While there is a reasonable amount of research on their health impacts, relatively little is known about the educational outcomes of these centres. Pandey (1991) suggests that the cognitive test scores achieved by children attending Anganwadi centres are somewhat higher than is the case for non-attenders; this study fails, however, to correct for heterogeneity across children which might arise through, for example, differences in innate ability or family background. Moreover it refers to the impact of the ICDS programme at a relatively early stage in its development. The present study evaluates the impact on subsequent educational participation of pre-school experience, whether this experience took place at an Anganwadi centre or elsewhere.

In many respects, the most obvious precursor to the present research is the work of Chamarbagwala (2008) who has studied the determinants of educational participation of 5-14 year olds in India with special reference to the role played, as explanatory variables, by rates of return to the various levels of education. Chamarbagwala estimates the model separately for children whose families are in the top and lower income groups, and finds that liquidity constraints affect participation in the expected direction. This finding is supported by the results reported in the present paper, although we use a somewhat different, reduced form, estimation method in which expected earnings do not feature explicitly as independent variables.

3. Data

Data for this study come from the 1997-8 Survey of Living Conditions in Uttar Pradesh and Bihar. This survey was conducted as part of the World Bank's Living Standards Measurement Study (LSMS) series, and the data are freely available from the LSMS website.[3] The LSMS was begun in 1980 as a project that would allow household surveys to be run in developing countries to gather data on social and economic outcomes. The data cover such topics as income and expenditure, economic activity, health, education, housing and utilities. Data are now available for over 30 countries, and these surveys have spawned several hundred research publications.[4]

In the case of India, the survey has been confined to the northern states of Bihar and Uttar Pradesh, both of which border Nepal and which form part of the cluster of BIMARU (Bihar, Madhya Pradesh, Rajasthan and Uttar Pradesh) states which are often referred to as being underdeveloped. Specifically, data were collected from villages located in the Eastern and Southern regions of Uttar Pradesh , and in the Northern and Central regions of Bihar. The two states are heavily dependent on agriculture, and both have poverty rates well above the national average. There has

[3] http://econ.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTRESEARCH/EXTLSMS/0,,contentMDK:21387345~isCURL:Y~pagePK:64168445~piPK:64168309~theSitePK:3358997,00.html

[4] Though remarkably few on India. But see, for example, Meenakshi and Ray (2002).

been some rapid industrialisation in the years since economic liberalisation, though much of this postdates the survey date.

With specific reference to education, Bihar is somewhat more developed than the Indian average, while Uttar Pradesh is considerably less so. Data on educational achievement for standard 3-5 pupils suggest that, in Uttar Pradesh, only 55.2 per cent of pupils can read in their own language (compared with an average for India of 66.4 per cent); only 42.8 per cent can pass a simple numeracy test involving subtraction (59.4 per cent for India as a whole), and only 11.1 per cent can read English (compared with a national average of 17.2 per cent).[5] The two states under consideration differ considerably also in terms of the proportion of pupils attending private schools: in Bihar some 7.4 per cent of pupils attend private schools, while in Uttar Pradesh the proportion is 29.1 per cent.[6]

The survey draws on some 2250 households (1035 in Bihar and 1215 in Uttar Pradesh). These households contain some 14493 individuals. Of these, 2311 are aged between 10 and 16 years inclusive, and these form the sample that is used in the analysis of the present paper. There are missing observations for 10 of the sample, so much of the statistical analysis is based on a sample size of 2301.

Descriptive statistics for key variables appear in Table 1. Of the 10-16 year olds in our sample, some 58 per cent are enrolled (*enroled*) in education at the time of survey. About 8 per cent of the sample ever attended pre-school (*presch*). A little over 90 per cent are Hindu (*hindu*), the remainder being Muslim. Around 18 per cent are from upper castes (*upper*). There are somewhat more boys than girls in the sample – 55 per cent are male (the *sex* variable taking unit value for males, and zero for females).

Several variables used in our analysis concern the household. We examine, for example, the role played by income per household member (*hhincpp*). This averages just over 10000 rupees per year, but comes with a very high dispersion. Around 20 per cent of households in which our sample reside receive remittances (remit) from donors living elsewhere. Household size (*hhsize*) averages about 8, but again there is considerable dispersion; the smallest household has 2 member, but the largest has 29.

There are also variables which are related to the village in which the households reside. In particular, we examine village facilities, specifically in the educational sphere. We have data on the existence of facilities in the village – secondary school, (*sec*), middle school (*mid*), primary school (*prim*), and Anganwadi centre (*angan*) – and the distance, in kilometres, of the nearest such facilities (*secdist*, *middist*, *primdist*, *angandst*).[7] As might be expected, the distance to the nearest facility tends to increase with the level of education, so that, on average, secondary schools are further away from their constituency than are primary schools. This is so partly because economies of scale for secondary schools, where specialist teachers and resources are required, suggest a minimum efficient scale that is larger than is the case for schools catering for lower levels of education.

[5] In Bihar, the corresponding figures are 68.8, 69.8 and 25.5 per cent respectively. These data come from http://pratham.org/aser07/ASER07.pdf.
[6] http://pratham.org/aser07/ASER07.pdf.
[7] The data on distance are grouped. We use midpoints, and 15km for the category representing the furthest distance.

4. Analysis

We examine the determinants of educational participation by estimating a logit model of participation for respondents aged 10-16 inclusive. The results appear in Table 2, and we focus initially on those in the first column.

The results suggest that being Hindu and coming from an upper caste, both of which may reflect social and economic advantage, raise the probability with which a sample member remains in education. Our finding on the importance of religion echoes the results of Unni (2007) who shows that the wage returns to education are lower for religious minority groups than for the Hindu majority. Unsurprisingly, as children age, they become less likely to remain in schooling.

Boys are much more likely to stay in education (long) than girls. The reasons underpinning this are likely to be several. Labour market discrimination and the roles played by the sexes in society may both contribute to a perception that education is a better investment in the case of boys than in that of girls. The gender disparity in education, specifically in the BIMARU states, has been the subject of comment by Kingdon (2005, 2007). She attributes much of the gender disparity to household fixed effects.

Household size exerts a positive influence on the probability with which a child remains (long) in school. This is an interesting finding in that we had no strong prior expectation on the direction of influence of this variable. Households that are large because they have many children are quite different from those that are large because they have many adults. In the former case, children represent competing demands on limited resources, and we might expect the size of the household to reduce the likely time spent by each child in school. In the latter case, on the other hand, the earnings of many adults may be used to support children through education. It would appear, therefore, that further information on household composition is needed in order to throw more light on these issues. Such an analysis is possible using the LSMS data, but has not been attempted here.

Children from households with a high level of income per capita are more likely to stay (long) in education than others. This is unsurprising as it implies that parents (and the wider family circle) who are more able to pay for their children's education (both through direct costs – tuition, books, uniforms and the like – and through the indirect cost represented by foregone earnings) do so. It implies also that there is potential for educational attainment to be raised in these regions by providing more support (guaranteed access to credit, perhaps?) targeted at low income households.

For much the same reason, children from households in receipt of remittances are more likely than others to remain (long) in education. These remittances may come from family members or other donors living elsewhere in India or abroad. Unfortunately, the survey does not allow further analysis of the source of remittances.

A further determinant of educational participation is the distance of the household from the nearest educational facilities. Distance to primary school is not significant, but distance to middle and (to a lesser extent) secondary schools is associated with a

fall-off in educational participation at these levels. Opportunity cost provides an obvious explanation.

All results reported above are based on logistic regression with controls for the industry in which the head of household, if salaried, works. These estimated coefficients on these controls are all insignificant with the exception of the miscellaneous services category (which has a significantly positive coefficient). This category includes workers in public administration, defence, health and sanitation, education and social services, recreation, culture, and personal services.

The final determinant of educational participation which we consider concerns the main theme of the paper – namely the impact of the early years experience of the child. We find that those children with experience of early years education are more likely than others to stay (long) in education. There is a possibility that this variable is endogenous, in the sense that the parents of children with the potential to go far in education might be more likely than others to invest in their pre-school education. We check for this by instrumenting pre-school experience using the presence of Anganwadi facilities in the village of residence as an instrument. The results are reported in the second column of Table 2. We find that the coefficient on *presch* increases markedly and remains significant; the remaining results are robust to the instrumentation procedure, confirming that pre-school education has a long lasting effect on the propensity of children to remain in schooling. The increased magnitude of the coefficient on *presch* suggests that the effects of endogeneity bias are dominated by those of recall bias. An alternative interpretation is that *angan* is an inadequate instrument; it is possible that Anganwadi centres opened in many villages during the ten years prior to the survey and that the existence of such centres in 1997-98 does not necessarily imply access to early years education when our sample of children were in the pre-school age group.

The central finding reported above has clear resonance with the recent American literature on the impact of early years education. The results are robust also to the correction of standard errors due to clustering of observations within households (Aitkin *et al*., 1981; Moulton, 1986); the corrected standard errors are shown in the third column of the table.

5. Conclusions

The results reported here provide some early indication that the beneficial impact of pre-school education is not confined to the United States. While recent studies in that country have focused on labour market outcomes, we have concentrated here on the impact that pre-school education has on subsequent participation in formal education. It appears that, after instrumenting for early years education, there is a significant positive benefit attached to pre-school in the form of longer duration in the education system.

There is clear scope for further work in this area. First, the present study, limited as it is by data considerations, is focused on only two states. National analysis would be welcome but must await the collation of more suitable data. Secondly, any study which relies on instrumentation to correct for endogeneity bias begs questions about

the validity of instruments. While the choice of instrument in this case seems to us to be natural and defensible, experimental methods clearly offer a superior research methodology. The difficulty is that the results of any experiment initiated now would be a long time in coming, as we would have to wait over ten years to know the impact of a pre-school experiment on subsequent educational choices. Thirdly, the issue that has been addressed in the United States – namely the labour market impact of early years education – has not been addressed here at all owing to data limitations. Longitudinal data collected over a lengthy period are highly desirable, but failing that it would be useful to have recall data on pre-school and other participation in education incorporated into microdata sets that investigate labour market outcomes.

Table 1 Descriptive Statistics

Variable	Mean	Standard deviation
enroled	0.5807	0.4936
presch	0.0852	0.2793
hindu	0.9022	0.2971
upper	0.1778	0.3825
hhsize	8.0541	3.8419
sex	0.5565	0.4969
age	12.7170	2.0863
remit	0.1965	0.3974
prim	0.8360	0.3704
mid	0.2683	0.4432
sec	0.1017	0.3023
angan	0.3323	0.4711
primdist	0.7700	0.7752
middist	2.6648	2.2980
secdist	4.7186	3.8349
angandist	0.4241	1.3629
hhincpp	10143.35	34150.37

Table 2 Results of logit analysis

Variable			
presch	0.5137	15.2730	15.2730
	(2.92)	(2.73)	(2.39)
hindu	0.8151	0.7333	0.7333
	(4.89)	(4.38)	(3.65)
upper	1.4493	1.4326	1.4326
	(9.91)	(9.80)	(7.88)
hhsize	0.0963	0.0963	0.0963
	(6.86)	(6.86)	(6.59)
sex	1.3943	1.3951	1.3951
	(14.04)	(14.05)	(13.27)
age	-0.2778	-0.2829	-0.2829
	(11.72)	(11.92)	(11.60)
remit	0.3087	0.2946	0.2946
	(2.50)	(2.39)	(2.00)
primdist	-0.0080	-0.0178	-0.1781
	(0.13)	(0.28)	(0.25)
middist	-0.0772	-0.0734	-0.0734
	(3.40)	(3.23)	(2.81)
secdist	-0.0172	-0.0154	-0.0154
	(1.25)	(1.12)	(0.98)
hhincpp	5.31×10^{-6}	5.01×10^{-6}	5.01×10^{-6}
	(3.30)	(3.12)	(2.88)
constant	1.4924	0.3701	0.3701
	(4.15)	(0.65)	(0.57)
n	2301	2301	2301
Log likelihood	-1297.43	-1298.05	-1298.05
Likelihood ratio $\chi^2(15)$	537.49	536.26	
Pseudo R^2	0.1716	0.1712	0.1712

Notes: z values in parentheses. Each model also includes a full set of single digit dummies for the industry in which the head of household is employed.

References

Aitkin, Murray, Anderson, Dorothy and Hinde, John, 1981. Statistical modelling of data on teaching styles, *Journal of the Royal Statistical Society A*, 144, 419-461.

Chamarbagwala, Rubiana (2008) Regional returns to education, child labour and schooling in India, *Journal of Development Studies*, 44, 233-257.

Cunha, Flavio, Heckman, James J., Lochner, Lance and Masterov, Dimitriy (2006) Interpreting the evidence on life cycle skill formation, in E. Hanushek and F. Welch (eds) *Handbook of the economics of education*, volume 1, Amsterdam: North Holland.

Cunha, Flavio and Heckman, James J. (2007) The technology of skill formation, *American Economic Review*, 97, 31-47.

Cunha, Flavio and Heckman, James J. (2008) Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation, *Journal of Human Resources*, forthcoming.

Currie, Janet and Thomas, Duncan (1995) Does Head Start make a difference?, *American Economic Review*, 85, 341-364.

Currie, Janet (2001) Early childhood intervention programs: what do we know?, *Journal of Economic Perspectives*, 15(2), 213-238.

Duraisamy, Palanigounder (2002) Changes in returns to education in India, 1983-94: by gender, age cohort and location, *Economics of Education Review*, 21, 609-622.

Heckman, James J. (2008) Schools, skills and synapses, NBER working paper 14064.

Heckman, James J., LaLonde, Robert J. and Smith, Jeffrey A. (1999) The economics and econometrics of active labor market programs, in O. Ashenfelter and D. Card (eds) *Handbook of Labor Economics*, volume 3A, Amsterdam: North Holland.

Heckman, James J. and Masterov, Dimutriy V. (2007) The productivity argument for investing in young children, *Review of Agricultural Economics*, 29, 446-493.

Heckman, James J. and Rubinstein, Yona (2001) The importance of noncognitive skills: lessons from the GED testing program, *American Economic Association Papers & Proceedings*, 91, 145-149.

Heyneman, Stephen P. (1980) Investment in Indian education: uneconomic?, World Development, 8, 145-163.

Karoly, Lynn A. and Bigelow, James H. (2005) *The economics of investing in universal preschool education in California*, Santa Monica: Rand.

Kingdon, Geeta G. (2005) Where has all the bias gone? Detecting gender bias in the intra-household allocation of educational expenditure in rural India, *Economic Development and Cultural Change*, 53, 409-452.

Kingdon, Geeta G. (2007) The progress of school education in India, *Oxford Review of Economic Policy*, 23(2), 168-195.

LaLonde, Robert J. (1986) Evaluating the econometric evaluations of training programs, *American Economic Review*, 76, 604-620.

Lynch, Robert G. (2007) *Enriching children, enriching the nation*, Washington: Economic Policy Institute.

Meenakshi, J.V. and Ray, Ranjan (2002) Impact of household size and family composition on poverty in rural India, *Journal of Policy Modelling*, 24, 539-559.

Moulton, Brent R., 1986. Random group effects and the precision of regression estimates. *Journal of Econometrics*, 32, 385-397.

Pandey, Hema (1991) Impact of preschool education component in Integrated Child Development Services program on the cognitive development of children, *Journal of Tropical Pediatrics*, 37, 235-239.

Unni, Jeemol (2007) Earnings and education among social groups, in A. Shariff(ed.) *State, markets and inequalities: human development in rural India*, New Delhi: Orient Longman.