Monograph:

Ahmad, U., McIntosh, S. orcid.org/0000-0002-9492-5455 and Popli, G. orcid.org/0000-0003-2919-2627 (2019) Selection and performance in post-compulsory education. Working Paper. Sheffield Economic Research Paper Series, 2019014 (2019014). Department of Economics, University of Sheffield ISSN 1749-8368

The University Of Sheffield.

Department Of Economics.

Sheffield Economic Research Paper Series.

Selection and Performance in Post-Complusory Education

Authors: Uzma Ahmad, Steven McIntosh, Gurleen Popli

ISSN 1749-8368

SERPS no. 2019014

June 2019

www.sheffield.ac.uk/economics

Selection and Performance in Post-Compulsory Education

Uzma Ahmad[a*], Steven McIntosh[b] & Gurleen Popli[b]

[a]* Corresponding Author: Sheffield International College, University of Sheffield, Sheffield, UK. uzma.ahmad@sheffield.ac.uk

[b] Department of Economics, University of Sheffield, 9 Mappin Street, Sheffield, S1 4DT, UK.

June 2019

Abstract

This paper investigates the determinants of participation and performance in post-compulsory education, controlling for the selection into post-compulsory education and prior attainment, using a unique primary dataset on pupils studying in the post-compulsory grade in 2011-2012 from one district of the Punjab province of Pakistan. The main findings of the paper show that participation and performance in post-compulsory education are two different processes, with participation being driven by availability of post-compulsory institutions within travel distance, while performance, once in post-compulsory education, is determined by ability. The results further highlight that distance reduces participation most for those living in rural areas.

JEL Codes: J24, I20

Keywords: post-compulsory education, participation, performance, distance, selection, Pakistan

Acknowledgements: We gratefully thank Nicolas Van de Sijpe and Pamela Lenton for helpful comments on an earlier draft. We would also like to thank the participants at the Fifth Lisbon Research Workshop on Economics, Statistics and Econometrics of Education, Lisbon, January 2019 for their comments.

1. Introduction

In this paper, we investigate the determinants of (1) the decision to participate into post-compulsory education (i.e. when continuing into education is a choice, usually made at the age of 16 years), and (2) performance in post-compulsory education, controlling for selection into post-compulsory education. In our analysis, along with other key drivers such as prior attainment and family socioeconomic status (SES), we specifically investigate the accessibility of post-compulsory education, measured as the distance to the nearest institution that provides post-compulsory education, as a determinant of participation. For our analysis, we use a unique primary dataset collected from Pakistan, with detailed information on individual and family characteristics, which is then linked with an administrative data to obtain educational outcomes.

There exists literature investigating the determinants of participation in post-compulsory education, and a limited separate literature investigating performance in post-compulsory education. However, there exist no studies to our knowledge, which investigate jointly the determinants of participation and performance in post-compulsory education.[1]

The often-discussed determinants of the decision to participate in post-compulsory education include prior attainment of the individuals, their family SES, and perceived benefits of schooling (Eckstein and Wolpin, 1999; Chevalier et al., 2013). Some studies also link the local labour market conditions with enrolment in post-compulsory education, with evidence suggesting that individuals are more likely to participate in education if the local labour market

[1] There are studies that investigate both participation and performance in education together. A large proportion of these studies, however, are in the context of Conditional Cash Transfer (CCT) programs (for a review of these studies see Fiszbein and Schady, 2009), which focus on compulsory education (mainly between the ages of 6 and 16 years). A few studies which are outside the CCT framework focus on primary education (for example, Duflo, 2001). Our paper is significantly different from the existing studies in the literature in two key ways: (1) in the CCT framework participation is mandatory and monitored as a part of the program; this is not the case in our study, where participation is a choice. (2) Unlike other studies in the literature, we focus on participation and performance in post-compulsory education.

is weak as education becomes more attractive due to a high local unemployment rate; for example, see McIntosh (2001) for Germany, Netherlands, Sweden and England, Bradley and Lenton (2007) for Britain, and Clark (2011) for England.

Similarly, factors determining performance in education, in general, include: ability, family SES, teacher quality and school characteristics; for example, see Chevalier et al. (2013) for the UK, Black et al. (2005a) for Norway, Engin-Demir (2009) for Turkey, Kasirye (2009) for Uganda, and Garcia (2014) for Russia. For post-compulsory education, specifically, the impact of working while in school has also been studied; a review by Neyt et al. (2017) finds that while working whilst in school has an impact on educational choices (engagement and continuation), it does not have an impact on education performance.

In developing countries there are additional determinants associated with educational participation and attainment. On the demand-side, poor and resource-constrained families see little immediate benefits of education, and on the supply-side the availability of, or access to, schools is limited. Accessibility of schools is an important consideration as costs associated with education increase with distance; these costs can be financial (travel and reallocation costs), temporal (commuting time) and psychological (inconvenience and unpleasantness of commuting every day). While CCTs have been a common policy response to the demand-side constraints; for the supply-side factors, the common policy response has been to build more schools or to reduce the distance costs to schools.

The existing literature looking at the impact of accessibility of schools on both participation and attainment largely focuses on the new schools built at the primary level (Duflo, 2001; Burde and Linden, 2013; Kazianga et al., 2013) and the reduction in the distance cost for access to middle schools (Muralidharan and Prakash, 2017).[2] There exists very little

[2] Of the two supply-side policies one could argue that building new schools, relative to reducing the distance costs, is likely to be costlier in terms of more qualified teachers and resources needed for institutions catering to higher levels of education. Further, distance costs are more amenable to change

evidence on the impact of accessibility to institutions providing post-compulsory education. One exception is Dickerson and McIntosh (2013), who are the first to study the impact of distance to an institution on the decision to participate in post-compulsory education at age 16. Using data from the UK, they find an overall small negative effect, with higher distance reducing participation in post-compulsory education.

Why Pakistan? Pakistan is a developing country facing difficult challenges in its education system, of which school dropout is one.[3] Despite the vast and growing literature examining educational outcomes in developing countries, the literature on educational outcomes in Pakistan is still in its infancy. Pakistan is an interesting case to study the impact of accessibility and availability of educational institutions on participation in post-compulsory education as in Pakistan parents have the flexibility to choose between private or government (state-funded) schools, according to their affordability. This choice is not innocuous, as it exacerbates the existing inequalities in the society, with individuals from relatively poor backgrounds accessing the free education provided by government schools, and the relatively rich accessing the private independent schools which charge fees.[4]

The present study uses primary survey data on 611 pupils from the Bahawalpur district of Punjab, Pakistan, studying in grade 12 (last year of schooling) at age 17-18 years,[5] to explore the determinants of performance and participation in post-compulsory education, controlling for previous attainment in grades 9 and 10, the last two years of compulsory education. A

relative to other long-term determinants of participation such as prior attainment and family background.

[3] Pakistan has the highest proportion of out-of-school children in South Asia, with roughly one in 10 of the world's primary-school-age children who are not in school living in Pakistan. A quarter of year 8 (lower secondary school, age 12 years) children drop out of education, with children in rural areas more likely to drop-out relative those in urban areas. Source: UNICEF (2013).

[4] In Pakistan in urban areas 61% of the school enrolment for compulsory education, age 6 to 16 years, is in private school; the corresponding figure for the rural areas is 21%. (Source: ASER, 2015)

[5] The ages written next to each grade, for Pakistan, are indicative, which assumes that every child is enrolled in formal schooling by the age of 6 years and progresses through the grades without repetition.

Heckman-selection model (Heckman, 1979) is used when analysing performance in grade 12 to control for potential sample selection bias due to unobserved characteristics being associated with the decision to leave (or continue in) education after achieving compulsory education. Proximity to nearest post-compulsory institution, measured as the distance to the nearest institution and living in an urban location, are used as instruments for selection into post-compulsory schooling, since both measure the accessibility and availability of post-compulsory educational institutions. The results show a lack of significant selection effects, which suggests that participation and performance are not jointly determined. Participation is determined by the accessibility of post-compulsory institutions, while performance in grade 12 is driven by ability, which is measured by previous attainment.

The present study contributes to the literature in the following ways. (1) It is the first study (to the best of our knowledge) to jointly investigate the determinants of participation and performance in post-compulsory education for both developed and developing countries. (2) It is the first study for a developing country, investigating the effects of proximity to nearest institution providing post-compulsory education. (3) It is the first study for Pakistan looking at the determinants of performance in higher grades, grade 12 taken at age 17-18, conditional on participation. (4) The present study takes into account the differences across regions (urban/rural) in the analysis of participation in Pakistan.

The remainder of the paper is structured as follows. Section 2 reviews the studies looking at accessibility of post-compulsory education from both developed and developing countries, and the relevant literature on performance and participation in post-compulsory education in Pakistan. Section 3 describes the data and discusses the key variables used in the analysis; section 4 presents the methodology employed, and summary statistics of variables used; section 5 discusses the empirical results; and section 6 concludes with some policy implications.

2. Literature review

2.1. Participation in post-compulsory education

There are studies which look at the effect of distance on participation in university education; Frenette (2006) for Canada; Spiess and Wrohlich (2010) for Germany; Gibbons and Vignoles (2012) for England. For both Canada and Germany evidence suggests that pupils living within commuting distance are more likely to attend university, relative to those who live out of commuting distance. Further, distance to the educational institution is found to matter more for those from low-income families. For England, results suggest little effect of geographical distance on the participation decision, but a strong effect on the choice of institutions.

In terms of the post-compulsory education participation decision, Dickerson and McIntosh (2013) are the first to investigate the effect of distance to the nearest educational institution on a pupil's immediate post compulsory participation decision in England. Their study highlights the small overall effect of the distance variable; with an additional kilometre in distance leading to a 1.5 percentage points decrease in the probability of participation in post compulsory academic study. In line with the literature the study finds a strong correlation between post compulsory participation, prior attainment and family background. They also find that distance affects women more than men. Further, distance has more significant impact on young people who are on the margin of participation in post compulsory education (according to their prior attainment and family background).

The Dickerson and McIntosh (2013) study is closest to our paper. Our paper, however, is different in two significant ways. First, we are different in the way we measure distance. Dickerson and McIntosh measure distance 'as a crow files' which could be substantially different from real distance, while we measure the 'actual' distance. The actual distance is measured by driving from pupils' homes to nearest educational institution using data on the location of the homes and schools. The latter arguably reflects the true distance cost of participation; none of the previous studies have measured distance using this method. Second, while Dickerson and McInstosh look only at participation, we look at both participation and performance in post-compulsory education.

2.2. Determinants of performance

There are a few papers which look at the effect of school accessibility and educational performance. Falch et al. (2013) look at the impact of travel time on the probability of graduation from upper secondary school in Norway. They find that reduced travel time has a positive effect on graduation, with the effect strongest for students with mediocre prior academic achievement i.e. those at the margin of graduation. Tiger et al. (2017) use data from Brazil and find a negative causal impact of commuting durations on the academic performance of sixth-grade students, after controlling for a range of family and school characteristics. Asahi (2014) looks at the effect of improved school accessibility on educational performance of students, using data from Santiago, Chile. Increased school accessibility is measured by a new subway line which reduced the distance between schools and their nearest subway station. The key finding of the paper suggests that increased accessibility is associated with substantially lower test scores, which the author explains by an increase in school enrolment (thus increasing the class size) for the more accessible schools.

2.3. Studies from Pakistan

Lloyd et al. (2009) look at factors responsible for primary and middle school dropout in the Punjab region of Pakistan. They use a longitudinal dataset, covering seven years from 1997-2004, which tracks changes in both the school and the family environment of individuals. Attending a government-funded primary school and the family experiencing an unwanted birth (used as an instrument to capture an exogenous household shock) increases the likelihood of dropout; while living in a better-off neighbourhood, mother having any education, and availability of post primary schooling reduce the chances of dropout. Results also suggest higher dropout rates for girls relative to boys.[6]

For Pakistan, much of the focus on performance in education has been on the learning differences observed across the different school types: private and government. Almost all studies (Alderman et al., 2001, Arif and Saqib, 2003, and Das et al., 2006, focussing on primary schools, and Aslam, 2009, looking at middle schools) find that the attainment of children in private schools is much higher relative to government schools, even after controlling for a range of household characteristics.[7]

Ours is one of the first papers for Pakistan to look at participation and performance in post-compulsory education. We also contribute to the limited literature on the impact of accessibility to educational institutions on participation at post-compulsory levels of education.

3. Data and empirical specification

This paper uses primary survey data collected by one of the authors. The data are for 611 pupils studying in grade 9, in 2008-09, in the Bahawalpur district of the Punjab province of Pakistan.[8] The data were collected using a random sampling technique over all private and government

---

[6] Low levels of literacy and gender gap in school enrolment resulted in the Government of Punjab, Pakistan, implementing a gender-targeted CCT (Female Secondary School Stipend (FSSS)) in 2004. The program was implemented in the 15 districts of Punjab, all of which had literacy levels below 40% in the population aged 10 years and older; Bahawalpur was one such district with a literacy rate of 35%. The program gave a stipend to families conditional on girls being enrolled in middle school (grades 6-8) in government-funded schools. The program had a positive impact on enrolment of girls in government schools (Chaudhury and Parajuli, 2010). While this CCT covered the area from which we obtain our data, the fact that it targeted only girls, in government-funded schools, for middle school grades means it should not have any bearing on our analysis.

[7] There is a significant gap in the learning levels of children in the private and government-funded schools. For urban areas 62% of the children in the last year of primary school (age 10 years) enrolled in private school could read a story in their native language (Urdu/Sindhi/Pashto), compared to 51% in government schools; corresponding figures for the rural areas are 67% and 52% respectively. (Source: ASER, 2015)

[8] Grade 9 in Pakistan is equivalent to grade 10 in the UK education system (a year before taking the GCSEs), when children are 14-15 years old. For details of the Pakistan education system see Appendix A of the paper.

schools, registered with the Education Department, in the district. Out of the 40 such schools, 15 school names were drawn, producing a sample of 6 government schools and 9 private schools. Each school was then approached to acquire their consent to participate in the study, all schools with the exception of one private school giving their consent, resulting in a sample of 14 schools: 6 government and 8 private. Among the 14 selected schools, all grade 9 pupils were given an opportunity to participate in the survey, a total of 700 pupils. For those pupils who were willing to participate, further consent was acquired from their parents. At every stage, schools, pupils and parents were assured of the confidentiality and anonymity of the survey.

The choice of Punjab and particularly Bahawalpur district for the field survey was based on the following grounds: (1) Literature on private schooling in Pakistan has noticed that much of the expansion in private schooling has been particularly prominent in the province of the Punjab (Aslam, 2009). (2) Bahawalpur is the largest district of Punjab in terms of area; it has a high number of reputable educational institutions, as well as a large number of government schools available in the city.

Data were collected from the pupils using pupil questionnaires filled in by the pupils themselves. Pupil questionnaires provided information on personal characteristics such as gender and age, as well as physical activity, hours of study at home, health status and mode of transport to school. A family questionnaire was sent to the home of each child and returned to the school authorities the next day completed by either parent (or the child completed the questionnaire by asking the parent questions if the parent was illiterate), containing information on parents' education, family size, family income, number of rooms, house tenure, etc.

Data on exam scores were subsequently obtained for sampled pupils studying in grade 9 in 2008-09, grade 10 in 2009-10, grade 11 in 2010-11, and grade 12 in 2011-12. Grades 9 and 10 are the last two years of compulsory schooling while grades 11 and 12 are post-compulsory schooling years. The exam scores used in this study are for high stake exams, as performance in these exams is considered by the universities and colleges in Pakistan at the time of giving admissions to higher education.

All schools in the sample are affiliated to the Board of Intermediate and Secondary Education (BISE) which is responsible for conducting exams in grades 9, 10, 11 and 12 in all government and private schools. The data on pupils are matched with administrative examination data held by the BISE, using unique pupil identifiers to obtain their test scores. Test scores from these exams are a reliable measure of students' academic performance as the examination system is uniform all over the Punjab province. The exam papers are the same across all schools and exams are held at the same time, scheduled by the BISE. The only difference is the medium of instruction and the language in which exams are set, where the Urdu language is used in all government schools and English language in the private schools. The exams are nationally set, so marking standards are the same and exams are marked outside the schools by the same set of external examiners, leaving little chance of systematic manipulations, tempering of results, or leakages of exam papers, while minimizing the incidence of cheating (Kingdon, 1996). Further, Clotfelter et al. (2010) argue that tests/exams being external to school as compared to within school means students are also more likely to take them seriously.

The variable 'distance to nearest post-compulsory education institution', in kilometres, was calculated using information on the actual home address of the pupils, recorded for all sampled students at the time of the first survey. The distance was measured by driving from the pupils' homes to the nearest feasible post-compulsory educational institution, not the school that they actually attended, the idea being to capture the accessibility of the nearest post-compulsory education institution for each pupil.

Participation is identified by whether pupils are observed in the administrative data with grade 12 results. It could be argued that the education dropout observed in the dataset could be in fact be due to missing data on individuals who are actually continuing to study, rather than to a decision not to participate in post-compulsory education.[9] One reason for missing data could be relocation of families out of the district. Using the unique pupil identifier, pupils can be searched for in two adjacent districts of Bahawalnagar and Rahimyarkhan, as the BISE Bahawalpur, is responsible for registration of students and conducting exams in all these three districts of Punjab. A search was undertaken for all missing pupils in the administrative data for the two adjacent districts. However, we found none of the pupils in the adjacent districts, so moves to at least these two districts can be ruled out. It could be possible that the families have moved to another district further away, where they can no longer be tracked; if this is the case then we have no way of tracing these families, however this seems unlikely as migration out of the Bahawalpur district is very limited.

After making every attempt to trace the students via the administrative data, we can make a reasonable claim that if we do not observe a student in the list of grade 12 exam results, it is almost certainly because they are no longer in education and not simply due to missing data. It is therefore people making a choice over whether or not to participate in post compulsory education.

3.1. Empirical specification

The main aims of this paper are to investigate the determinants of (1) participation in post-compulsory education, and (2) performance in post-compulsory education, controlling for selection into post-compulsory education. Correcting for selection here is important, as the pupils who dropout from education are more likely to be those with low ability, who are less

[9] Missing observations due to sample attrition is not possible in our sample, as all information used here, other than the test scores, is collected in the first wave, subsequently we rely only on the administrative data for test scores. It is also possible that students attended grades 11 and 12 but did not take the school leaving exam at the end of grade 12, which is recorded in the administrative data. We still classify them as non-participants.

motivated, from low quality schools and from disadvantaged home backgrounds (Alivernini and Lucidi, 2011; Chowdry et al., 2013).

Let d_i^* be the unobserved utility, derived by individual i, from participation in post-compulsory education. Assuming a linear in parameter specification, this utility depends on a set of covariates, given by vector z_i, such that we have: $d_i^* = \gamma' z_i + \varepsilon_i$, where γ is the vector of parameters to be estimated and ε_i is the random error, which we assume to be normally distributed. We do not observe d_i^*, rather what we observe is a dummy variable indicating participation:

$$P(d_i = 1|z_i) = P(d_i^* > 0|z_i) = P(\varepsilon_i > -\gamma' z_i) = 1 - F(\gamma' z_i) \quad (1)$$

Equation (1) is the participation equation, where $d_i = 1$ if the pupil graduated from grade 12 i.e. finished post-compulsory education; this is estimated as a probit model.

Vector z_i includes a range of covariates such as: prior attainment of the pupil, measured as the total score in grade 9; measures of pupil attributes in grade 9, such as age, health problems, gender and birth order; a set of variables capturing the family characteristics of the pupil measured in grade 9, such as fathers' education, mothers' education, fathers' income, whether or not parents own their house; as well as a dummy variable capturing the type of school attended (private or government school) by pupils in grade 9. The participation model has two additional variables: distance to the nearest post-compulsory education institution, and a dummy variable capturing whether the pupil lives in an urban or a rural area.

Table 1 gives the education participation rate of the pupils in our survey. Starting with the 611 pupils from the randomly drawn schools in grade 9, we were able to match 591 (96%) of them with the administrative data and obtain their grade 9 test scores. Of these 591 students, by the end of grade 12, 325 (55%) graduate. There appear to be two dropout points from education; some are not completing grade 10 (or not doing the exams at least) while others are not participating in grade 11/12. We define the post-compulsory education participation variable as all those who were observed in administrative data in grade 9 and then are also observed in the same administrative data in grade 12 i.e. they participate in post-compulsory education and take the school leaving exam at the end of grade 12. There is just one pupil who did participate after the end of compulsory education, in grade 11, but did not take the school leaving exam at end of grade 12, we drop this pupil from our analysis. The group of non-participators in post-compulsory education include a group who dropout after grade 9 and a group who dropout after grade 10. In our main analysis we do not distinguish between the two dropout points. However, it is possible that there are two separate decisions that need to be modelled, i.e. decision 1 – whether or not to take the grade 10 exams, and decision 2 – whether or not to participate in post-compulsory schooling and take the grade 12 exams. To check the validity of treating the two decisions as a single decision, as a robustness check we separately consider the second decision in isolation, modelling the selection into post-compulsory education conditional on having completed the grade 10 exams.

To investigate the pupils' performance at the post compulsory level we use a value added education production function (Hanushek and Rivkin, 2012). In this specification, we use both current and past educational inputs, and take into account the effect of prior attainment, which captures the effect of previous teachers and schools, and individual ability The empirical specification we estimate is given as:

$$Y12_i = \varphi Y9_i + \beta' x_i + \delta\lambda_i + u_i \tag{2}$$

where $Y12_i$ is the academic performance measured by grade 12 total score of the i^{th} pupil, and $Y9_i$ is the past performance of the same pupil, with parameter φ capturing the effect of past performance.[10] In our analysis we use grade 9 score as the prior achievement. It could be argued

[10] An Education Production Function (EPF) is used to examine the relationship between the test scores and a range of input variables. In an ideal situation the input variables would capture all the past and present characteristics of the student, and their family, school and teachers; further, we should be able to distinguish these inputs from the inheritable endowments. However, given data limitations, where all the desired information is rarely available to the researcher, certain assumptions are made to estimate

that it would be more natural to control for grade 10 scores, since grade 10 is the last year of compulsory education. However, as Table 1 shows, relying on grade 10 scores to provide the measure of prior attainment would involve a significant reduction in the number of available observations, due to some pupils not taking their end of compulsory schooling exams. The main analysis is therefore undertaken using grade 9 scores, though we do undertake analysis to ensure that the results are robust to using grade 10 scores.

The vector x_i is the vector of the pupil's individual and family characteristics, containing all variables mentioned in the participation model, except for the distance variable and the urban dummy, and with the addition of variables indicating type of institution attended in grade 12. Information for the type of school attended at post-compulsory level is identified from the administrative data. β is the parameter vector to be estimated, while u_i is the normally distributed error term.

Not every pupil is observed in grade 12, as we cannot assume participation is random, we correct for selection using the Heckman selection approach, with equation (1) serving as the first stage of the analysis. λ_i is the inverse Mills ratio from the first stage regression, this is used as an additional regressor in equation (2). δ is the associated parameter taken as a measure of sample selection bias; if δ is statistically different from zero, the null hypothesis of no sample selection bias is rejected.

3.2. Excluded variables

For identification of the selection process, we need at least one variable to satisfy the exclusion restriction, that is, a variable that directly affects participation, but has no impact on performance once in post-compulsory education. In our specification, we consider two such

---

the EPF. Todd and Wolpin (2003) discuss the full set of assumptions underlying the different specifications of EPF. In our analysis we estimate the value added specification of EPF, where it is assumed that the past test score captures all the historical home and school inputs, as well as inherited endowments for which we do not have data. The value-added specification further assumes a certain degree of persistence in learning (i.e. $\varphi > 0$) between grades.

variables: distance to the nearest post-compulsory educational institution and an urban/rural dummy. The basic intuition for using distance as an instrument for participation, is that it captures the costs of attending a relevant academic institution (Dickerson and McIntosh, 2013), which can be particularly constraining for those from disadvantaged background.

Actual distance, in kilometres, is not used often due to the complex data requirement. In lieu of distance, some studies use a rural/urban dummy in analysing participation in university or post-secondary education, such as Kane and Spizman (1994) for the US, and Christofides et al. (2001) for Canada. These studies, however, do not draw direct inferences on distance, they implicitly assume that youths living in urban areas are closer to education institutions than rural youths and hence are more likely to participate in education. There is a criticism that distance and urban/rural location should not be used interchangeably; for example, Frenette (2006) argues that there could be rural areas which are better served by educational institutions, while for some small urban area accessibility might be limited (though Frenette was talking about universities and not post-compulsory educational institutions). In the current analysis we use both distance and rural/urban location as separate controls in the selection equation while studying the impact of geographical constraints on the participation decision. Use of the urban/rural dummy also helps us address another criticism, that it might be the case that the distance is higher but the government transport is better in urban areas.[11]. We therefore also consider an interaction term between the two variables.

3.3. Descriptive statistics

A description of the variables used in this study is given in Table 2, and the descriptive statistics for these variables are provided in Table 3. The grade 12 score is the sum of scores over seven

[11] One could argue that there are differential resources between the two (urban and rural) locations that affect performance, such as type of institution. To address this concern we use a separate control for type of institution attended in grade 12 in the model. However, we acknowledge that this is only one potential difference in resources, and there still could be other remaining differences between the two types of area.

subjects; four subjects are compulsory (Language, English, Religion and Pakistan Studies) and three are optional depending on the field of study chosen. In total five fields of study exist in grade 12: premedical group, pre-engineering group, general science group, humanities group, and commerce group. The total score for each subject is 200 except Religion and Pakistan Studies, these two subjects having a total score of 75 each. The maximum total marks in grade 12 are therefore 1150. Prior attainment (grade 9 score) is the overall/aggregate score in seven subjects being offered in grade 9 in schools; of these four subjects are compulsory (Language, English, Religion and Mathematics) while three subjects are elective, differing across the two possible fields of study: Science and Arts. The maximum total score in grade 9 is 625. In our analysis we standardize all test scores to have a mean of 0 and standard deviation of 1.

There are three categories for type of institution in grade 12: appearing in the exam after attending private school, after attending government school, or as an independent candidate. Independent candidates are those who do not attend any formal institution; usually these are rural youths, females, or disabled individuals. They undertake self-study and preparation for exams, and send their exam admissions independently to their respective registered board. There are certain subjects/disciplines that cannot be taken as an independent candidate, particularly science subjects such as physics, chemistry, and biology. As these subjects have practical exams along with theoretical exams, attending regular classes is a requirement for them. In the current dataset, 43% of pupils in grade 12 attended a private institute, 46% a government institute and 11% of pupils appeared as independent candidates. In grade 9 there are only two possible categories of institution: private and government school. The option of being an independent candidate does not exist in grade 9. In grade 9, 32% of pupils are from private schools with the remainder from government schools.

The mode of transport variable captures the transport used daily to get to the school attended in grade 9; this variable is used to proxy the level of poverty. The schools are not located in pupils' postcode areas. Therefore, pupils from relatively well-off families can use a car, bus or motorcycle to reach schools while pupils from relatively poorer families either use cycles or walk to schools. In our sample 42% of the pupils use a car or bus.

The average age of students in grade 9 is 15 years. The minimum age to start primary education in Pakistan is 5 years; however, there is no enforcement of this, and parents can send their children late to school, particularly in rural areas. About 11% of the pupils did not report their age. Rather than dropping these individuals from the sample, we use a dummy variable indicating 'age missing'. Among those who do report their age there is a huge variation, with the maximum reported age being 20 years and the minimum reported age being 10 years.[12]

We also control for birth order of the pupil under consideration, as there is evidence to suggest that higher birth order has a negative impact on the educational performance of children (Black et al., 2005b). For health problems, the question asked of the pupils is 'do you have any chronic health problems affecting your studies?'; 28% of the sample report such problems. Studying time at home is also important as many students have to help with household chores, especially girls, which can negatively affect their performance at school. Half of the respondents are female. The other family variables in our analysis include ownership of house (78% of the households own their own home), number of siblings, and mother and father's education, and father's income.

[12] The age distribution in the data is as follows:

Age (years)	missing	10	11	12	13	14	15	16	17	18	19	20
Number of observations	60	1	2	5	23	107	161	156	32	10	8	1

Age in our sample is self-reported, which means there is a chance of misreporting. 79% of our sample falls within the age range of 13 to 16 years, which can be assumed to be a reasonable age range for a grade 9 student. Comparable figures for all of Pakistan are 82% in the rural areas and 87% in the Urban areas (ASER, 2015). As a robustness check analysis was repeated by dropping all those who do not report age and those with age below 14 years. The results (not reported, but available on request) are qualitatively the same as those reported below.

The variable distance to nearest post-compulsory institution, measured in kilometres, is the actual measured distance from pupils' homes to the nearest post-compulsory education institution; this is not the actual school attended in grade 12. The mean distance to the nearest post-compulsory education institution is 8 kilometres. To account for non-linearity, log of distance is used in this study (Newbold and Brown, 2015). Urban is a dummy variable taking value 1 if the pupil is living in an urban area and 0 if she is living in a rural area.[13]

4. Results and discussions

4.1. Determinants of participation in post-compulsory education

Table 4, column (1) presents the probit analysis of equation (1) showing determinants of participation in post-compulsory education. The dependent variable is measured as a binary variable, taking value of 1 if a pupil participated in a post-compulsory education (grade 12), and a value of 0 otherwise. The column reports the average marginal effects with robust standard errors.

Prior attainment has no significant effect on participation in post-compulsory education. This result is different from previous findings for developed countries (McIntosh, 2001). Attending a private school in grade 9 significantly decreases the likelihood of participation in post-compulsory education by 7 percentage points holding other things constant. Analysis of the raw data show that pupils in private schools in grade 9 are actually far more likely to participate in post-compulsory education (as we might expect). The negative coefficient observed here in the multivariate regression is due to the controls used, with distance being the key variable affecting the private school coefficient. The data show that pupils who go to private schools live much closer to their nearest institution, which benefits their participation, and because distance is a key determinant of participation, then when we control for distance

[13] The definition of rural and urban areas for Bahawalpur are taken from the following Asian Development Bank report: https://www.adb.org/sites/default/files/project-document/67386/26479-pak-pcr.pdf

and hold it constant (and so take away the private schools' advantage), the private school coefficient becomes negative.

The mode of transport variable is positively and significantly associated with the probability of participation in post-compulsory education, using a bus/car as a mode of transport being associated with a 5 percentage point higher likelihood of participation in post-compulsory education. Mode of transport is used as an indicator of disadvantage. The estimated coefficient therefore indicates that a lower level of disadvantage has a positive association with participation in post-compulsory education.

Among individual characteristics, the only statistically significant coefficient is time spent studying at home and birth order. Studying at home more than two hours is associated with a 6 percentage point lower probability of participating in post-compulsory education; this could be due to those struggling in their studies, needing to spend more time studying at home to catch-up. Higher birth order increases the probability of participation in post-compulsory education.

Among family characteristics the only significant coefficient is father's income which is negatively and significantly associated with the likelihood of participation in post-compulsory education. The negative coefficient on fathers' income is surprising. In the raw data, father's income is higher for those who continue into education relative to those who do not. The negative coefficient observed here in the multivariate regression is therefore due to fact that other characteristics of the household are being controlled for, with income having no positive effect over and above that of these other characteristics.

The significant and negative coefficient on log of distance suggests that the further a pupil lives from a post-compulsory institution, the less likely he/she is to participate in post-compulsory education. A unit increase in log of distance reduces the likelihood of participation by 5 percentage points. The coefficient on the urban dummy is positive and significant, indicating that pupils living in urban areas are more likely to participate in post-compulsory education, living in an urban area increasing the probability of participation in post-compulsory education by 25 percentage points relative to living in a rural area. This is in line with the general hypothesis that urban areas potentially have more accessible institutions.

Our results suggest that past performance is not associated with participation in post-compulsory education, which is contrary to the findings often found in the literature, as discussed earlier. For our sample, participation in post-compulsory education is mainly determined by the distance to the nearest post-compulsory school and living in an urban area. This is an interesting finding as both variables measure the accessibility to post-compulsory education for young people and suggests that if institutions are made more accessible, participation in post-compulsory education can be increased. These results are consistent with the findings from previous studies (Dickerson and McIntosh, 2013; Frenette, 2006; Spiess and Wrohlich, 2010).

4.2. Determinants of performance in post-compulsory education

Determinants of performance without selection correction

For performance in post-compulsory education (grade 12) we start by estimating equation (2) using OLS, i.e. without correcting for selection effects. Results are reported in column (2) of Table 4. The dependent variable in this model is the standardized pupils score in grade 12.

The variable most strongly associated with performance in post-compulsory education is prior attainment, as measured by performance in grade 9; a one standard deviation (SD) increase in attainment in grade 9 is associated with a 0.8 SD increase in post-compulsory marks. This finding is consistent with previous literature such as Todd and Wolpin (2007) who find positive effects of past educational performance on current educational outcomes.

Attending a government school relative to a private school in the post-compulsory years decreases performance on the exam by 0.46 SD of marks. Similarly, appearing in the exam as an independent candidate, not through a private institution, is associated with a decrease of 0.35 SD of marks in the post-compulsory grade. These findings are consistent with previous literature. Both Aslam (2009) and Kingdon (1996) compared the relative efficiency of private and government schools and found that the achievement of students in private schools is higher, relative to students in government schools. This could be due to poor resources, including higher pupil to teacher ratios, in the government schools.

All other variables in this model have statistically insignificant coefficients.

Determinants of performance correcting for selection bias

Column (3) of Table 4 reports the results for performance from Heckman's specification, i.e. after correcting for selection (first stage for which is reported in column (1)). There is no significant difference in the estimated coefficients between the OLS and selection corrected specifications. The Wald test that all coefficients in the model (except the constant) are 0 is rejected.

The selection correction term (lambda) is insignificant; further, we cannot reject the null hypothesis that $\rho = 0$, where ρ (rho) is the correlation between the error terms of the first stage and the second stage regressions. This suggests that there is no selection problem, performance once in post-compulsory education is mainly determined by past performance and characteristics of the educational institution.[14]

4.3. Validity of exclusion restrictions

The excluded variables determining participation should not affect the outcome variable, performance in post-compulsory education. It is, however, possible to make a theoretical

---

[14] For both the OLS model and the Heckman specification (for both the first and the second stage estimates) we added the different categories of explanatory variables sequentially. Starting with the most conservative specification of only past performance in the performance equation and past performance and the two excluded variables in the participation equation, from this start other sets of covariates were added sequentially, until all the covariates were included (the specification presented in Table 4). The results presented and discussed in the paper remain unchanged both qualitatively and quantitatively. Results are available on request from the authors.

argument for the potential effect of the excluded variables on performance. For example, it might be the case that time spent on commuting directly reduces the time available for school work, thus reducing achievement. Also, the different neighborhood conditions in urban areas might influence achievement directly. For example, rural areas are less connected and have lower population density. Pupils in rural areas are therefore likely to have fewer peers as examples continuing in education, be more likely to have to work in the household or on the land, have poorer, less well resourced schools etc. Rural/urban status in particular might therefore be expected to pick up factors associated with school performance, and therefore be wrongly excluded from the attainment equation. To confirm that this is in fact not the case, the instruments were entered, one at a time, into the performance equation, to see if they have significant coefficients. First, we include only log of distance as an instrument in the participation equation and include the urban dummy in the performance equation; results are reported in column (1) of Table 5. Second, we include the urban dummy as an instrument in the participation equation and include log of distance in the performance equation; results are reported in column (2) of Table 5.[15] In both cases, the coefficient on the variable included in the first stage remains statistically significant, while the coefficient on the variable included in the second stage is found to be statistically insignificant. Neither distance nor rural/urban status appears to be associated with Year 12 attainment, conditional on the other control variables included. Further, rho and lambda in the second stage remain statistically insignificant indicating no selection.

One other potential concern would be that where people live is endogenous, i.e. they have specifically chosen their house location for the access it gives to schools. This could

[15] In each case, it is assumed that the single identifying instrument remaining is sufficient to identify the selection equation, so that the attainment equation is still estimated consistently, and so the coefficient on the variable added to the attainment equation provides an appropriate indication of its association with attainment.

potentially affect the instruments for reasons that might also affect participation, for example if those who value education more choose to live near schools. Distance, in this case, would not be a random exogenous variable but would have been endogenously determined. Evidence suggests that this does happen in the UK (Allen et al., 2010). For example, most school places are allocated on the basis of the catchment area, so that places are usually offered first to children who live nearest to a school (Gibbons, 2012). This is less relevant here as such decision-making is not usual in Pakistan. First, there is no official policy to assign school places on the basis of catchment area as there are no defined catchment areas or postcodes. Parents can choose any school depending mainly upon their affordability and then preferences. Secondly, families in Pakistan do not choose their house and move location for schooling reasons as it is hard for them to afford relocation. This potential criticism is therefore not particularly relevant in the existing case.

5. Robustness checks

In this section several robustness checks are applied to check the sensitivity of the results.

5.1. Changing sample to those observed in grade 10

One potential criticism of the main analysis is that there could be a further selection issue as there is still one more year of compulsory schooling involved after grade 9, from which some individuals drop out. The analysis is therefore redone, focussing on the post-compulsory participation choices of only those who made it to the end of grade 10 and sat the final exams in compulsory schooling. The results for this are reported in columns (1) and (2) of Table A1 in the Appendix.

The usable sample when using grade 10 as the potential dropout point is 419 observations. The results are similar to those already reported – both for the first stage i.e. participation, and the second stage i.e. performance. Similar results for the grade 9 and grade 10 samples indicate that the two decisions (completing compulsory schooling and proceeding to post-compulsory education) do not need to be modelled separately, but rather we can treat the 'complete grade 10' decision and 'enter grade 11' decision as a single decision as to whether to participate in post-compulsory education or not, with those choosing not to do the grade 10 exams simply dropping out a little earlier as they know that they will not be continuing to the post-compulsory level.[16]

5.2. Distance to the nearest school of the same kind

It could be argued that it is not the distance to the nearest school that matters, but the distance to the nearest school of the type the family would consider. For example, it could be argued that if a family would only consider going to a private school, then the fact that a government school is only 1 km away would be irrelevant, since it would be how close the nearest private school is that would be relevant. To address this criticism, we restrict our sample to the individuals where the measured distance to the nearest post-compulsory educational institution is of the same type as the one attended in the compulsory phase. For example, if a student in grade 9 was going to a private school, and the nearest post-compulsory institution (for which we have the distance) is a private school then we retain this individual in the sample; on the other hand if the nearest post-compulsory institution is a government school then we drop this individual from the sample. To do this we assume that pupils going to private (government) schools in grade 9 would prefer to continue in a private (government) post-compulsory institution. Once we restrict the sample in this way, we are left with 379 observations. The results for this are reported in columns (3) and (4) of Table A1 in the Appendix. The results using this restricted sample again remain qualitatively the same.

[16] All robustness checks reported below were repeated for the grade 10 sample. The results, not reported, remain qualitatively the same. Results are available on request from the authors.

5.3. Interaction of the excluded variables

It could be argued that distance has more impact on participation for those from rural areas where transport links are likely to be not as good. Individuals who live in rural areas typically have less access to public transport, a larger distance to travel, and a greater level of disadvantage. In the current dataset, the average distance to the nearest school for pupils from urban areas is 2 kilometres, while it is 16 kilometres for pupils living in rural areas. As a consequence, the participation of individuals living in rural areas may be more affected by the distance to the nearest school. To test this hypothesis, we estimate the participation equation with interaction effects between distance to nearest post-compulsory educational institution and the urban dummy. The results are reported in Table 6. The estimated coefficients for other control variables remain qualitatively the same as in Table 4, therefore they are not reported but are available on request. Distance to the institution remains a significant deterrent to participation. The coefficient on the urban dummy is now insignificant which suggests that if pupils live at zero distance from a school (literally next door to a school) then it does not make any difference to their participation whether they live in an urban area or rural area - the school is right there for them. The interaction term has a positive and a significant coefficient, suggesting that the negative effect of distance is much larger in rural areas. The second stage results, not reported but available on request, remain unchanged.

6. Conclusions

This paper fills the gap in the existing literature by exploring the joint determinants of participation and performance in post-compulsory education, controlling for any selection effects. The paper focuses on the impact of accessibility of institutions on participation in education. Accessibility is captured by two different variables: distance from pupils' homes to the nearest post-compulsory educational institution, and the location where pupils live – urban or rural area. For our analysis we use a unique survey data set on pupils studying in grade 9 at age 14-15 years from the Bahawalpur district of Punjab, Pakistan in 2008-9. The survey collects detailed information on the individual and family characteristics, and is then linked with administrative data to obtain exams scores. The distance variable is created using actual travel distance by driving from pupils' homes to the nearest post-compulsory educational institution.

The findings of the paper suggest that participation in post-compulsory education in Bahawalpur is strongly associated with distance to the nearest post-compulsory educational institution and urban location, while past performance, as an indicator to some extent of ability, is not found to be associated with participation. To investigate determinants of post-compulsory performance, a Heckman selection model is used, where distance and urban location are used as excluded variables that explain participation but not performance. The key variables associated with performance are past attainment and type of institution attended. We find no evidence of selection bias, suggesting that participation decisions and final performance outcomes are two completely different processes. Thus, those participating in post-compulsory education are not necessarily the most able, but those with the best access to post-compulsory education.

These results have a number of implications, with the efficiency of the post-compulsory education system in Pakistan being key amongst them. The results suggest that the young people most capable of doing well in post-compulsory education are not necessarily those who are participating. Given that prior attainment in the compulsory phase is shown to be a strong predictor of achievement in grade 12, participation by those with easy access, but not necessarily high ability, will affect the aggregate human capital being produced by the system. Particularly in a country such as Pakistan where the opportunity cost of funding for education is likely to be high, the consequences of such misallocation are clear.

The implications for policy are numerous. First, access to post-compulsory education should be restricted to those with the sufficiently high prior attainment to suggest a high likelihood of success. At present, although post-compulsory colleges typically offer places on the basis of prior attainment, there is no fixed pass mark for acceptance, no guidelines from education policy and colleges set their own minimum level of acceptance in order to ensure places are filled, depending on the numbers applying and what their marks are. Private colleges in particular have more freedom to set their admissions criteria, with ability to pay fees being one consideration.

Second, more help needs to be given to those who do not continue into post-compulsory education but have the ability and prior attainment to do well, had they done so. Lack of access to educational institutions has been highlighted in this paper as one possible deterrent factor, but there could be others in terms of family background and the need for financial support, or simply more information and guidance in communities where an absence of peers going into post-compulsory education means a lack of role models to inspire young people to do likewise.

Finally, focussing particularly on the access issue, though results from the analysis suggest that lower distance increases participation, the reason for this distance effect is not clear. It could be either financial cost or time cost due to frequent long travelling. If distance reduces participation due to longer travel time that involves longer commuting, then reducing the travel time by increasing the frequency of government transport is an effective policy. If distance reduces participation due to financial issues, then giving a subsidy, grant or loan to young pupils would be an appropriate option. For example, the Canada Student Loan Program (CSLP) provides loans to poor pupils (Frenette, 2006). Further, the results find that participation affected by distance is more a rural phenomenon than urban, with distance reducing participation more for those living in rural areas. Provision of more further-education institutions in rural areas and targeting rural pupils for further-education subsidies would be an optimal policy.

Given the unique nature of the data set allowing us to analyse participation and performance in detail, one drawback is that the sample size is small. Further, we would be cautious before generalising the findings to other parts of Pakistan, or indeed to other countries. For-example, the Bahawalpur district is an agricultural state in south Punjab, Pakistan, while other metropolitan areas may exhibit different patterns of participation or performance. Nevertheless, for this region, the results have provided useful information to policy-makers and educators on the separate processes driving participation and performance in post-compulsory education.

References

ASER Pakistan. Annual Status of Education Report Pakistan (2015).

Alderman, H., Orazem, P. F., & Paterno, E. M. (2001). School quality, school cost, and the government/private school choices of low-income households in Pakistan. Journal of Human Resources, 304-326.

Alivernini, F., & Lucidi, F. (2011). Relationship Between Social Context, Self-Efficacy, Motivation, Academic Achievement, and Intention to Drop Out of High School: A Longitudinal Study. The Journal of Educational Research, 104(4), 241-252.

Allen, R., Burgess, S., & Key, T. (2010). Choosing secondary school by moving house: school quality and the formation of neighbourhoods: The Centre for Market and Government Organisation working paper no. 10/238.

Arif, G.M., & Saqib, N. (2003). Production of cognitive life skills in government, private, and NGO schools in Pakistan. Pakistan Development Review, 42(1), 1–28.

Asahi, K. (2014). The Impact of Better School Accessibility on Student outcomes: Discussion Paper No. 0156, Spatial Economics Research Centre, London School of Economics.

Aslam, M. (2009). The relative effectiveness of government and private schools in Pakistan: are girls worse off? Education Economics, 17(3), 329-354.

Black, S. E., Devereux, P. J., & Salvanes, K. G. (2005a). Why the Apple Doesn't Fall Far: Understanding Intergenerational Transmission of Human Capital. The American Economic Review, 95(1), 437-449.

Black, S. E., Devereux, P. J., & Salvanes, K. G. (2005b). The more the merrier? The effect of family size and birth order on children's education. The Quarterly Journal of Economics, 120(2), 669-700.

Bradley, S., & Lenton, P. (2007). Dropping out of post-compulsory education in the UK: an analysis of determinants and outcomes. Journal of Population Economics, 20(2), 299-328.

Burde, D., & Linden, L. L. (2013). Bringing Education to Afghan Girls: A Randomized Controlled Trial of Village-Based Schools. American Economic Journal: Applied Economics, 5 (3): 27-40.

Chaudhury, N., & Parajuli, D. (2010). Conditional cash transfers and female schooling: The impact of the female school stipend programme on government school enrolments in Punjab, Pakistan. Applied Economics, 42 (28), 3565–83.

Chevalier, A., Harmon, C., O' Sullivan, V., & Walker, I. (2013). The impact of parental income and education on the schooling of their children. IZA Journal of Labor Economics, 2(1), 1-22.

Chowdry, H., Crawford, C., Dearden, L., Goodman, A., & Vignoles, A. (2013). Widening participation in higher education: analysis using linked administrative data. Journal of the Royal Statistical Society: Series A (Statistics in Society), 176(2), 431-457.

Christofides, L. N., & Hoy, M. (2001). Family income and postsecondary education in Canada. Canadian Journal of Higher Education, 31(1), 177-208.

Clark, D. (2011). Do Recessions Keep Students in School? The Impact of Youth Unemployment on Enrolment in Post-compulsory Education in England. Economica, 78(311), 523-545.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2010). Teacher credentials and student achievement in high school a cross-subject analysis with student fixed effects. Journal of human Resources, 45(3), 655-681.

Das, J., Pandey, P. & Zajonc, T. (2006). Learning levels and gaps in Pakistan. World Bank Policy Research Working Paper no 4067, The World Bank, Washington DC.

Dickerson, A., & McIntosh, S. (2013). The impact of distance to nearest education institution on the post-compulsory education participation decision. Urban Studies, 50(4), 742-758.

Duflo, E. (2001). Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment. American Economic Review, 91 (4), 795-813.

Eckstein, Z., & Wolpin, K. I. (1999). Why youths drop out of high school: The impact of preferences, opportunities, and abilities. Econometrica, 67(6), 1295-1339.

Engin-Demir, C. (2009). Factors influencing the academic achievement of the Turkish urban poor. International Journal of Educational Development, 29(1), 17-29.

Falch, T., Lujala, P., & Strom, B. (2013). Geographical Constraints and Educational Attainment. Regional Science and Urban Economics, 43(1), 164–176.

Frenette, M. (2006). Too far to go on? Distance to school and university participation. Education Economics, 14(1), 31-58.

Fiszbein, A., & Schady, N. (2009). Conditional Cash Transfers: Reducing Present and Future Poverty. Washington, DC: World Bank.

García, E. D. T. (2014). Determinants of educational outcomes: Analysis of the Republic of Tatarstan. Communist and Post-Communist Studies, 47(1), 39-47.

Gibbons, S. (2012). Big ideas: valuing schooling through house prices. Centrepiece, 17(2), 2-5.

Gibbons, S., & Vignoles, A. (2012). Geography, choice and participation in higher education in England. Regional science and urban economics, 42(1), 98-113.

Hanushek, E. A., & Rivkin, S. G. (2006). Teacher quality. Handbook of the Economics of Education, 2, 1051-1078.

Heckman, J. (1979). Sample Selection Bias as a Specification Error. Econometrica, 47(1), 153-161.

Kane, J., & Spizman, L. M. (1994). Race, Financial Aid Awards and college Attendance. American Journal of Economics and Sociology, 53(1), 85-96.

Kasirye, I. (2009). Determinants of learning achievement in Uganda. Economic Policy Research Centre, Uganda.

Kazianga, H., Levy, D., Linden, L. L., & Sloan, M. (2013). The Effects of 'Girl-Friendly' Schools: Evidence from the BRIGHT School Construction Program in Burkina Faso. American Economic Journal: Applied Economics, 5 (3): 41-62.

Kingdon, G. (1996). The Quality and Efficiency of Private and Government Education: A Case-Study of Urban India. Oxford Bulletin of Economics and Statistics, 58(1), 57-82.

Lloyd, C. B., Mete, C., & Grant, M. J. (2009). The implications of changing educational and family circumstances for children's grade progression in rural Pakistan: 1997–2004. Economics of Education Review, 28(1), 152-160.

McIntosh, S. (2001). The demand for post-compulsory education in four European countries. Education Economics, 9(1), 69-90.

Muralidharan, K., & Prakash, N. (2017). Cycling to school: increasing secondary school enrollment for girls in India. American Economic Journal: Applied Economics, 9(3), 321-50.

Neyt, B., Omey, E., Baert, S., & Verhaest, D. (2017). Does Student Work Really Affect Educational Outcomes? A Review of the Literature. IZA Discussion Paper No. 11023.

Newbold, K. B., & Brown, W. M. (2015). The Urban–Rural Gap in University Attendance: Determinants of University Participation Among Canadian Youth. Journal of Regional Science, 55(4), 585-608.

Spiess, C. K., & Wrohlich, K. (2010). Does distance determine who attends a university in Germany? Economics of Education Review, 29(3), 470-479.

Tigre, R., Sampaio, B., & Menezes, T. (2017). The Impact of Commuting Time on Youth's School Performance. Journal of Regional Science, 57(1), 28-47.

Todd, P. E., & Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. The Economic Journal, 113(485), F3-F33.

Todd, P. E., & Wolpin, K. I. (2007). The production of cognitive achievement in children: Home, school, and racial test score gaps. Journal of Human capital, 1(1), 91-136.

UNICEF (2013). Out-of-School Children in the Balochistan, Khyber Pakhtunkhwa, Punjab and Sindh Provinces of Pakistan. (https://www.unicef.org/pakistan/OSC_UNICEF_Annual_Report.pdf).

Table 1: Post-compulsory participation rate in the sample

	Respondents	Percentage of original sample	Education participation
Initial Sample	611	-	-
Academic progress recorded in grade 9	591	96%	
Academic progress recorded in grade 10	441	72%	75%
Academic progress recorded in grade 11	326	53%	55%
Academic progress recorded in grade 12	325	53%	55%

Notes: The above table tracks the participation in education of individuals over a period of time from grade 9 to grade 12, collected through a survey and later recorded through the matched administrative dataset from BISE, Bahawalpur using a unique pupil identifier. The last column shows the percentage of pupils who participate in education from within the sample of 591 pupils for whom we have grade 9 test scores from the administrative data set.

Table 2: Description of Variables

Variables	Description
Score in grade 12 (dependent variable)	Total marks obtained by pupil in grade 12
Prior attainment	Academic total score of grade 9
School variables	
Type of school in grade 12 (Reference category: private school)	1 if pupil attended private school, 2 for government school and 3 if appeared in exam as an independent candidate not through any school
Private school grade 9	Dummy variable equals 1 if pupil attended private school in grade 9 and 0 for government school
Mode of transport	Dummy variable equals 1 if pupil uses car or bus to reach school and 0, if he/she walks
Individual characteristics	
Pupil age	Age of pupil measured in years
Male	Dummy variable equals 1 if male and 0 for female
Birth order	Pupil's order standing among other siblings
Health problems	Dummy variable equals 1 if pupil has a chronic disease
Private coaching (tuition)	Dummy variable equals 1 if pupil takes private coaching other than school
Study hours per day at home	Dummy variable equals 1 if pupil studies at home more than 2 hours per day
Family characteristics	
Own house	Dummy variable equals 1 if parents own a house
Number of siblings	Number of brothers and sisters pupil has
Fathers' education	Father's education measured in years
Mothers' education	Mother's education measured in years
Father income per month	Income of father measured in Pakistani Rupees (PRs)
Exclusion Variables	
Distance	Distance measured in km, from home to nearest secondary ischool, not the school which they actually attended
Urban	Dummy variable (0/1) equals 1 if pupil is living in urban area and 0 for rural area

Notes: Other than score in grade 12 and the type of school attended in grade 12, all other variables are for grade 9.

Table 3: Summary statistics of variables

Variables	Obs.[1]	Mean	S.D.	Min.	Max.
Standardized score in grade 12[2]	304	0.00	1.00	-2.42	2.33
Prior attainment (standardized score in grade 9)[2]	566	0.00	1.00	-2.25	2.33
School variables					
Government school grade 12	304	0.43	0.50	0	1
Private school grade 12	304	0.46	0.50	0	1
Independent candidate grade 12	304	0.11	0.31	0	1
Private school grade 9	566	0.32	0.47	0	1
Mode of transport	566	0.42	0.49	0	1
Individual characteristics					
Pupil age (Years)	566	15.20	1.26	10	20
Pupil age missing[3]	566	0.11	0.31	0	1
Male	566	0.49	0.50	0	1
Birth order	566	3.03	1.83	1	11
Health problems	566	0.28	0.45	0	1
Private coaching (Tuition)	566	0.58	0.49	0	1
Study hours per day at home	566	0.75	0.43	0	1
Family characteristics					
Own house	566	0.78	0.42	0	1
Number of siblings	566	4.66	2.03	0	16
Fathers' education (Years)	566	10.10	5.14	0	20
Mothers' education (Years)	566	6.62	5.81	0	18
Fathers' income per month (PRs.)[4]	566	19254.61	46304.79	1000	1000000
Log of father's income per month	566	9.30	0.97	6.91	13.82
Exclusion Variables					
Distance (kilometres)	566	8.25	8.12	0.5	40
Urban	566	0.57	0.50	0	1

Notes:

1. The number of observations here are different to those in Table 1 due to usable sample.
2. Test scores have been standardized to have a mean 0 and standard deviation of 1.
3. Where pupil's age is missing, instead of dropping the pupil from analysis we create a dummy to capture that age is missing.
4. To get a relevance the mean income would be equal to £140 per month approximately; this is using the exchange rate as it was at the time of the survey. In today's (June 2019) exchange rate this would be £98.

Table 4: Participation in post-compulsory education (first stage, probit model), and determinants of performance in post-compulsory education (OLS coefficients, and second stage selection-corrected coefficients)

	(1)	(2)	(3)
Dependent variable:	**Participation**	**Score in grade 12**	
Variables	**Marginal Effects**	**OLS coefficients**	**Selection corrected coefficients**
Prior attainment (grade 9)	0.01	0.80***	0.80***
	(0.01)	(0.06)	(0.06)
School variables			
Private school, grade 9	-0.07**		
	(0.03)		
Government school, grade 12		-0.46***	-0.45***
		(0.09)	(0.09)
Independent candidate, grade 12		-0.35**	-0.34**
		(0.16)	(0.15)
Mode of transport	0.05***	-0.12	-0.12
	(0.02)	(0.08)	(0.08)
Individual characteristics			
Age	0.01	-0.04	-0.04
	(0.01)	(0.03)	(0.03)
Age missing	0.10	-0.53	-0.54
	(0.11)	(0.52)	(0.50)
Male	0.01	-0.04	-0.05
	(0.02)	(0.09)	(0.09)
Birth order	0.01*	-0.01	-0.01
	(0.00)	(0.02)	(0.02)
Health problems	-0.03	-0.05	-0.06
	(0.02)	(0.08)	(0.08)
Private coaching	0.00	-0.10	-0.09
	(0.02)	(0.07)	(0.07)
Study time at home	-0.06***	0.13	0.13
	(0.02)	(0.08)	(0.09)
Family characteristics			
Own house	-0.01	-0.01	-0.02
	(0.02)	(0.09)	(0.08)
Number of siblings	-0.01	0.03	0.03
	(0.00)	(0.03)	(0.03)
Fathers' education	0.003	-0.01	-0.01
	(0.00)	(0.01)	(0.01)
Mothers' education	-0.003	0.01	0.01
	(0.00)	(0.01)	(0.01)
Fathers' log income	-0.02*	0.06	0.05
	(0.01)	(0.05)	(0.06)

Dependent variable: Variables	(1) Participation Marginal Effects	(2) Score in grade 12 OLS coefficients	(3) Selection corrected coefficients
Excluded variables			
Log of distance	-0.05***		
	(0.02)		
Urban dummy	0.25***		
	(0.05)		
Constant		-0.09	0.01
		(0.69)	(0.68)
Lambda		-	0.15
		-	(0.24)
Rho		-	0.26
			(0.40)
Observations	566	304	566
R-square		0.661	
Wald Chi2 (Prob > ch2)[1]			541.42 (0.000)
Wald test of independency of equations i.e. Rho = 0, Chi2 (Prob > Ch2)[2]			0.37 (0.541)

Notes: *** p<0.01, ** p<0.05, * p<0.1. (.) report robust standard errors.

1. This is the Wald test that all coefficients in the regression model (except the constant) are 0.
2. This is the Wald test for the independence of the two equations.

Table 5: Check for exclusion restriction

Panel A: First stage, probit model for participation		
	(1)	**(2)**
Covariates	**Marginal effects**	**Marginal effects**
Log of distance	-0.204***	-
	(0.019)	
Urban dummy	-	0.354***
		(0.045)
Other controls[1]	YES	YES
Panel B: Second stage, determinants of performance in post-compulsory education		
Covariates	**Coefficients**	**Coefficients**
Log of distance	-	-0.026
		(0.078)
Urban dummy	-0.193	-
	(0.756)	
Other controls[1]	YES	YES
Lambda	-0.004	0.175
	(0.134)	(0.238)
Rho	-0.006	0.300
	(0.231)	(0.401)
Observations	566	566
Wald Chi2 (Prob > ch2)	595.52 (0.00)	536.79 (0.00)
Wald test of independency of equations i.e. Rho = 0, Chi2 (Prob > Ch2)	0 (0.979)	0.49 (0.483)

Notes: *** p<0.01, ** p<0.05, * p<0.1. (.) report robust standard errors.

1. The specification controls for all other variables as in Table 4.

See notes to Table 4.

Table 6: Interaction between the excluded variables. First stage results, marginal effects from probit model for participation

Covariates	**Interaction between (log of distance and urban dummy)**
Log of distance	-0.175***
	(0.057)
Urban dummy	-0.069
	(0.124)
Log of distance ***** urban dummy	0.133**
	(0.057)
Other controls[1]	YES
Number of observations	566

Notes: *** p<0.01, ** p<0.05, * p<0.1. (.) report robust standard errors.

The specification controls for all other variables as in Table 4.

Appendix

Brief Overview of the Formal Education System in Pakistan

The education system in Pakistan is divided into five different levels.

1) Primary Level (grades 1 to 5): ages 6 to 10 years
2) Middle Level (grades 6 to 8): ages 11 to 13 years
3) Secondary Level (grades 9 and 10): ages 14 to 16 years
4) Higher Secondary Level (grades 11 and 12): ages 17 to 18 years
5) University Level (leading to undergraduate and graduate degrees): age 18 years and above.

According to Article 25-A of the Constitution of Pakistan, the state must provide free and compulsory education from age 6 to 16 years, i.e. for grades 1 to 10. The ages written next to each grade are indicative, which assumes that every child is enrolled in formal schooling by the age of 6 years and progresses through the grades without repetition.

The Secondary School Certificate (SSC) includes grades 9 and 10 examinations, and is equivalent to GCSE/O level in England, while the Higher Secondary School Certificate (HSC) includes grades 11 and 12 examinations, this is equivalent to GCE A Level in England. Both SSC and HSC exams are conducted through the Board of Intermediate and Secondary Education (BISE). There are also alternative qualifications available in Pakistan, where SSC and HSC are replaced by Ordinary Level (O Level) and Advanced Level (A Level) which are managed by the British Examination Boards of Cambridge University, however the present study does not include these alternatives.

There are two main types of schools in Pakistan, private and government schools. Although similar in terms of their educational structure, they both differ in terms of finances and regulations. Most of the cost of operating the government schools is borne by the public exchequer. Private schools are owned by sole proprietors, trusts or non-governmental organizations (NGOs) or other forms of management; they work for profit and follow either the national curriculum or a curriculum approved by foreign educational institutions.

All government schools provide free education from grade 1 to grade 10, including free books and uniforms, while private schools charge fees ranging from 500 to 12,000 Pakistani Rupees (PRs) per month, depending on factors like their standard, reputation and location; further, there are separate charge for examinations, books, and uniform. For post-compulsory education, both private and government schools charge fees. Government schools charge fees up to 8,000 PRs, while private schools' fees vary between 15,000 and 150,000 PRs. Most private schools are co-educational while all government schools are single sex. In government schools, female students are always taught by female teachers and male students by male teachers, while in private school, it could be male or female teachers depending upon the availability. Some private schools involve entry tests at the time of admission but government schools do not.

All government schools are registered with the Education Department. Although every private school is required to be registered with the Education Department, most of these schools remain unregistered and consequently the size of the private school sector is unmeasured in government statistics (Aslam, 2009). All schools, private or government, which are registered with the Education Department, are linked with the administrative data, from which the test scores of the pupils can be obtained (similar to the National Pupil Database (NPD) in the UK). If a private school is not registered with the Education Department then they cannot be linked with the administrative data. In the sample used in this paper, there are pupils from both the private and government schools, and all the private schools are registered with the Education Department, so these can be linked up with the administrative data.

Table A1: Participation in post-compulsory education (first stage, probit model), and determinants of performance in post-compulsory education (OLS coefficients, and second stage selection-corrected coefficients) for different sample.

	(1)	(2)	(3)	(4)
	Sample restricted to those observed in grade 10		**Sample restricted to those for whom the nearest school at grade 12 is the same as in grade 9**	
Dependent variable:	**Participation**	**Score in grade 12**	**Participation**	**Score in grade 12**
Variables	**ME**	**Coefficients**	**ME**	**Coefficients**
Prior attainment (grade 9)			0.00	0.91***
			(0.02)	(0.07)
Prior attainment (grade 10)	-0.01	0.76***		
	(0.01)	(0.07)		
School variables				
Private school, grade 9	-0.06**		-0.08*	
	(0.03)		(0.04)	
Government school, grade 12		-0.41***		-0.44***
		(0.10)		(0.10)
Independent candidate, grade 12		-0.42**		-0.27
		(0.19)		(0.22)
Mode of transport	0.06***	-0.04	0.08**	-0.03
	(0.02)	(0.08)	(0.03)	(0.10)
Individual characteristics				
Age	0.004	-0.02	0.02*	-0.05
	(0.01)	(0.04)	(0.01)	(0.04)
Age missing	0.04	-0.20	0.27*	-0.54
	(0.10)	(0.57)	(0.16)	(0.54)
Male	-0.01	-0.10	0.02	-0.09
	(0.02)	(0.10)	(0.03)	(0.11)
Birth order	0.02**	-0.001	0.01	-0.03
	(0.01)	(0.03)	(0.01)	(0.03)
Health problems	-0.03	-0.08	-0.03	0.04
	(0.02)	(0.08)	(0.03)	(0.09)
Private coaching	-0.02	-0.18**	0.01	0.01
	(0.02)	(0.07)	(0.02)	(0.08)
Study time at home	-0.06**	0.07	-0.10***	0.08
	(0.03)	(0.09)	(0.04)	(0.11)
Family characteristics				
Own house	-0.03**	0.001	-0.03	0.04
	(0.02)	(0.09)	(0.03)	(0.11)
Number of siblings	-0.01**	0.05*	-0.01	0.04
	(0.01)	(0.03)	(0.01)	(0.03)
Fathers' education	0.003	-0.002	0.00	-0.01
	(0.002)	(0.01)	(0.00)	(0.01)
Mothers' education	-0.003	-0.001	-0.00	0.01
	(0.003)	(0.01)	(0.00)	(0.01)

	(1)	(2)	(3)	(4)
	Sample restricted to those observed in grade 10		**Sample restricted to those for whom the nearest school at grade 12 is the same as in grade 9**	
Dependent variable:	**Participation**	**Score in grade 12**	**Participation**	**Score in grade 12**
Variables	**ME**	**Coefficients**	**ME**	**Coefficients**
Fathers' log income	-0.01	0.04	-0.03	-0.02
	(0.01)	(0.06)	(0.02)	(0.07)
Excluded variables				
Log of distance	-0.04**		-0.08***	
	(0.02)		(0.03)	
Urban dummy	0.26***		0.33***	
	(0.06)		(0.08)	
Constant		-0.07		0.55
		(0.80)		(0.85)
Lambda		0.220		0.325
		(0.176)		(0.269)
Rho		0.371		0.587
		(0.292)		(0.542)
Observations	419	419	379	379
Wald Chi2 (Prob > ch2)		507.41		588.78
		(0.00)		(0.00)
Wald test of independency of equations i.e. Rho = 0, Chi2 (Prob > Ch2)		1.32		0.95
		(0.250)		(0.329)

Notes: *** p<0.01, ** p<0.05, * p<0.1. (.) report robust standard errors.
See notes to Table 4.