*Full Length Research Paper*

# Expectation returns and households' decision in the schooling of their children in Khyber Pakhtunkhwa Pakistan

## Shahzad KHAN[1]*, Puangkaew LURHATHAIOPATH[2] and Shusuke MATSUSHITA[2]

[1]School of Life and Environmental Sciences, University of Tsukuba, Ibaraki-ken Japan.
[2]Faculty of Life and Environmental Sciences, University of Tsukuba Japan.

This study aims to study households' expectations for their children's academic performance in Khyber Pakhtunkhwa, Pakistan. Education has a significant role in increasing the productivity and income level of an individual in a society. Household education, income, distance from school, gender discrimination within household and cost of education can affect parents' expectations of their children. The present study aims to understand why households in Khyber Pakhtunkhwa, Pakistan do not send their children to school, when free and compulsory public schools are available. A two stage Propensity Score Matching approach was applied in the study. Data were estimated through two stages Propensity Score Method. The first stage consists of probability model, used to estimate the propensity score of the characteristics of household. In the second stage, each household group was matched to predict households with similar propensity score values. Literacy rate had a negative effect on the completion of school for household members under 20 years. This means that most school graduates are unable to even read and understand. This is due to dynamic causes, like untrained teachers, lack of facilities, old syllabus, and poor quality education and socio-economic background of the household. But the main delinquency is the failure of educational policies attributed to inadequate economic structure and political instability since the existence of Pakistan. Pakistan government and policy makers need to take initiatives to improve the socio-economic condition of the individuals in the province, and also to give awareness about the importance of education in the region. Further investigation is needed to know the effect of other heterogeneous treatment.

**Key words:** School completion, propensity score matching, Khyber Pakhtunkhwa.

## INTRODUCTION

Education plays a significant part in the economic development of a country. Pakistan is a developing country situated in the Western part of South Asia. The total population of Pakistan is 188 million; 62% of its population lives in rural areas, and 60% depend on agriculture for their livelihood. The gross domestic product (GDP) of Pakistan is 243.6 billion US$, and the GDP per capita income of the people is 1316.14 US

Dollars (World Bank, 2016). At national poverty line, the poverty head count ratio is 29.50%, and its inflation rate is 8.04%, which is difficult for the people who face extreme poverty (UNESCO, 2016; World Bank, 2016).

According to education for all (EFA) 2008 report, India has high chance to achieve the Universal Primary Education of Millennium Development Goals (MDG) by year 2015, while Pakistan cannot. This is because without the effort of families, communities and policy makers it is not possible to achieve the target of 88% literacy rate by the year 2015. For gender and rural and urban disparities, it is 58% which is far behind the universal primary goal achievements (ASER, 2016).

According to the study of Barro and Lee (2001), the average year of schooling in 1960 to 2000 has significantly increased in adult men and woman in South Asia except Afghanistan, whereas Bangladesh and Nepal are particularly unassertive in the region. The increasing literacy rate of India is comparatively better than that in Pakistan with its economy growing faster in recent years. Sri Lanka has the best literacy achievements in South Asia with 92.63% literacy rate, which is preeminent in the region, compared to India and Pakistan. In case of Gender Parity Index, there is a huge gender gap in the average educational attainment in South Asian countries such as India, Pakistan, Bangladesh and Afghanistan (World Bank Data, 2016).

The education system of Pakistan consists of subsequent structures which are pre-primary, primary, middle (lower secondary), secondary, higher secondary, and higher education (University Level of Education). Pre-primary school children's ages are 3 to 5 years; primary consists of 1st class to 5th class with children aged 5 to 9 years. Middle school (lower secondary) includes 6th class to 8th class, and the ages of the children are 10 to 12, while secondary level consists of class 9 to 10th. Some diploma and vocational schools give admission after the secondary class certification (UNESCO, 2012). Pakistan is far behind in achieving its target in universal primary education despite its policy commitment and assurance by the government of Pakistan.

Currently, Pakistan gross enrolment rate is 85.9% while the specified goal was to achieve 100% by the year 2015. 21.4 million children aged 5 to 9 years are in primary school, while 68% are already enrolled in schools in which 6.5 million are girls (44%) and 8.2 million are boys (56%) (EFA Report, 2014). Though the Government of Pakistan provides free primary education, there are still children that do not go to school, because their parents are illiterate. Cost of education is a big problem, which makes households hesitate to send their children to school (Kadzamira and Rose, 2003).

The investigation from developing countries, including India and Pakistan, showed that only 65% of grade 3 students were able to solve a one-digit subtraction, while only 59% were able to solve simple multiplication problem and only 24% were able to read and write. Also, less than 20% could understand a simple paragraph of Urdu (Pritchett and Beatty, 2015).

Therefore, concentrating on interventions, curricula and policies, it is important to conduct a fruitful investigation to solve the problems faced by the school children. Studies in such educational research are significant for policy makers, researchers, teachers, parents and administrators (Graesser, 2009).

Education is an investment in human capital development. Productive and highly skilled labor force is the result of systematic reforms in educational policies and is the requirement of time, which shows the quality of a system. This mechanism is based on the implementation of planning and good policy formulation (Hallak, 1995).

Since Pakistan came into being, a total of nine booklets have been published on educational policies, in which only one document 1972 was established, while all the remaining eight documents failed to dwell on public welfare. This is due to improper funding, political ambiguity and flaws in the administrative structure of the country. The funds which were granted by the international organizations to improve absenteeism in schools were not properly utilized by the government, hence it badly affected the education sector in the last few decades (Khan et al., 2016).

The quality and standard of education in rural areas is dropping and causing huge rural/urban disparities and inequalities in Pakistan. The Education for All Report (2013 to 2014) stated the educational status of low income countries across rural areas of Pakistan; there is widespread learning crisis due to deficiencies in quality of education (Agrawal, 2014).

However, the public-private strategy of education in Punjab is adopted by the government but there is no proper way to understand the gap in educational achievements. The unsatisfactory level of students' achievements in Punjab, Pakistan indicates that many children are unable to pass the test in their learning levels (Andrabi et al., 2007).

Pakistan is lagging behind in attaining her goals in the field of free and compulsory education. This research will elaborate the key problems in the country that hinder households to send their children to schools.

## Objectives of the study

1. To identify the factors that diversify households' interest in educating their children and their future expectations.
2. To know the relationship that exists between households' background and their participation in their children's education.
3. To classify the central factors that hinder households' children from being educated.

This research article will be much helpful to policy makers, and the researchers in finding out the hindrances to educational achievements in Pakistan, and household

behavior towards their children's education.

## Prior theories related to returns to education and dynamic factors

The research can generally be fragmented into three primary areas of empirical research, though of course there are many overlaps in terms of sub- groupings and practice. The three principal areas are as follows: theories connected to living discourses, households' assets, education, distance from school to home and participation of the child in his education.

The theories related to children's discrimination and bias in relation to education, and finally theories related to provision and use of basic services provided to household and their children such as free education services in Pakistan were studied. There is limited literature on theoretical clarification as to why diverse individual households have different thinking and expectations from education. Different research works have only focused on certain schools instead of analyzing and going deeper into why households' decision is important in the education of their children.

Patrinos and Psacharopoulos (2002) in their research estimate that 27% of average global private return is due to primary education. They also indicate in their research that the contributions of primary education to better natural resource management play an important role in growing the economy of a country.

Becker (1962) and Mincer (1974) evolution in human capital is based on education and training, which has direct relationship with the earnings of an individual. The coefficient of school years in Mincerian earning function points to the returns to schooling, as an additional year in school increases the earnings of an individual.

Hanushek and Kimko (2000) described that education is a circulation of technological information in an economy. Studies from Pakistan suggested low rate of returns to education at different levels of education, paralleled to other emerging economies. The earning function for different levels of education was applied for the estimation of results (Haque, 1977; Guisinger et al., 1984).

Epstein and Jansorn (2004) stated that household participation has an adjacent association with the success of schools as well as in the development of students. Those schools who provide high quality environment involve the parents of the students in direct communication about their children's future expectations.

Parents, teachers and students' participation can improve the educational level of a student in school as well as the child's interest level. Lloyd and Grant (2009) stated that the quality of education in school depends on the engagement of parents in the school activities. He further stated that, it is clear that participation of parents with even little educational background contributes to children's academic progress.

In addition, even those households who are educated up to 10th grade take interest in the education of their children. However, families in Pakistan focus more on boys' future earnings and goals than girls because they leave their parents after marriage; on the other hand, parents have more expectations from boys than girls in terms of finances and better life style (Zeira and Dekel, 2005). Mansory (2007) described the main causes of school dropout in Afghanistan. He mentioned that due to early marriages, boys and girls leave school and start to work because of household responsibilities. With the unpaid domestic work at home, when girls after marriage shift to their husbands' house they stop school and concentrate on their husbands and other family members. Family background is also an important factor and one of the main causes of school dropout. Odaga and Heneveld (1995) described in their research that households consider girls' education as a waste of money since parents think that girls are to be married as early and as soon as possible.

Sathar and Lloyd (1994) noted distance as a hindrance for most of the female students, resulting in high level of dropout or long absenteeism. If the distance from home to school is more than a kilometer, most of the girls lose interest to go to school, and this is due to the poor infrastructure in the rural areas as well as the cultural barriers. There are no high gender differences in primary schools in Lahore, Pakistan; whereas, in other parts of Pakistan, the distance from girls schools to home has significant negative impact on their enrolments (Alderman et al., 1996).

Another study in India has shown that using bicycle for long distance school plays an important role in increasing education attainment. Muralidharan and Prakash (2013) described that the enrollment rate in rural India has increased as girls are given bicycle to school, which also reduces the gender gap in the region. They stated that there was 32% increase in secondary school enrolment rate, and 40% decrease in the gender gap.

Gertler and Glewwe (1990) stated that distance from schools, local teachers and teaching quality influence students' interest in their studies. The indirect cost of communication and transportation system is also a big problem for the parents, who send their children to school from their out-of-pocket charges. Research from Asia offers many comprehensive studies on household decision for their children, and expectations for future goals. Studies also found out that there are number of problems associated with the wellbeing of children in some parts of Asia, and this is a big concern for the educational system.

## Household expectations and returns from education

Households' decision-making and the return expectation

in case of gender are different. From the assumption of household model, we can observe that parents are altruistic in their comportment towards their children. Parents in terms of human capital accumulation and consumption care about the present and future expectations (Emerson and Souza, 2002a).

It is also discussed in some research studies that parents have special preference and sympathies with the same gender, like father spends more time during work with his son and mother with her daughters while doing work at home (Thomas and Perry, 1994).

According to the study of Emerson and Souza (2002), domestic work at home has a negative impact on the schooling of the female child. The girl child mostly spends her time at home and helps her mother in household chores, while the father has the same relationship with his son. They further explained in a study in Brazil that parents who do not participate in the schooling of their female children tend to give more attention to their boys' education. There is gender discrimination in here (Emerson and Sousa, 2002b).

Horowitz and Souza (2016) using the instrumental variable approach, described the robust monotonically decrease association among the instrumented income of households and the progression of educational attainment of households. This association depends on the child's academic performance in poor household, which is an important issue for policy implications. Shah and Anwar (2014) stated in their research work in Southern Punjab Pakistan, that parental education and family income have a significant effect on the education attainment of their children. They found that parental participation in academic activities motivates their children, thus improving their cognitive skills and academic achievements.

Munda and Odebero (2014) explained that EFA is still a big challenge for the poor households of the developing countries, that are unable to finance their children's education. They discussed in their findings that there is a significant positive association between unit cost and academic attainments.

Despite the financial aid given by government for education, poor households in developing countries still find it difficult to send their children to school. In another study, Karemesi (2010) found out that examination fee, cost of text books, school uniforms, transportation, sports and feeding are a big problem towards achieving the Universal Basic Education, especially for the low income families. Literature in various field mentioned the dynamic reasons for the low participation of household in the education of their children. This is the basis for low literacy rate and high dropout, and all due to the poor erection of educational policies that affect the poor household to achieve their education in developing countries.

In Pakistan, very little attention has been given to this problem, and there are very limited discussions on households' interest in educating their children. This research will enhance the weaknesses of households in Pakistan. The study will explore some of the issues which stop households to send their children to school. This research will also provide an opportunity to policy makers, organizers and authorities in the field of education to make policies and strategies in the light of the growth and development of the society. The study will also explore the literature regarding the behavior of household towards the education of their children.

## METHODS OF ESTIMATION AND DATA

To obtain better response for research questions, it is important to use mixed approach to understand the problems and expand the thoughtful research hitches, especially in the field of social sciences (Creswell, 2013).

Many researchers have debated about the use of qualitative and quantitative methods in research. In qualitative method, researchers use phenomenological approaches, which are naturalistic inquiries; while in quantitative research, experimental and non-experimental quantitative approaches are used to measure hypothetical questions. Quantitative research is based on casual determination, generalization and prediction of findings (Patton, 1990).

In this study, we used the quantitative method in order to approach the flaws in the educational sector of Pakistan and to understand the main factors that affect the schooling decision of households. This research evaluates the impact of public educational policy implication on households' expectations from returns to education in Province Khyber Pakhtunkhwa, Pakistan.

The main purpose of the counterfactual evaluations is to elaborate what would have happened if the policy had not been taken. The quasi-experimental approach is used to answer this question in order to pretest and compare the treated and control groups. Therefore, in a non-random research, it is important to estimate the matching method by using statistical techniques. The detailed discussion for problem evaluation is mentioned below.

### Problem evaluation

Empirical methods used in development economics have been technologically advanced to give answers to counterfactual questions, as studies endeavor to estimate the mean effect of the treatment group participating in the program.

An inference is required to know about the outcome for the treatment group; and when they are not treatment group, it is called control group. The experimental methods have advantages over non experimental studies, which have the capability to create a control group with the characteristics of the same dissemination as the treatment groups. For such methods, the difference of mean outcome will be calculated as treatment effect. With respect to their participation, the status and characteristics of treated and control groups are different. A biased result occurs between the two groups, when estimating the treatment effect as the difference of mean outcomes.

In order to calculate the average effect of an individual program in non-experimental method, matching method is generally used. By using this method, we compare the outcomes of individual groups who participated with non-participants, and observed characteristics are chosen on the basis of similarity in matches. Let us assume that we have two groups of household members: those who enroll in the school year (2013 to 2014) and completed their education and those who did not complete their education in the same year. Differentiating these two groups was done based on participation status.

**Conceptual framework and assumptions**

The important issue in evaluating the impact of education on household behavior is the specification of the average treatment effect. Rosenbaum and Rubin (1983) defined the average treatment effect $(\Delta_i)$ in a counterfactual framework as;

$$\Delta_i = Y_S - Y_N \tag{1}$$

$Y_S$ is the outcome condition on schooling $(P=1)$
$Y_N$ is the outcome condition on non-schooling $(P=0)$

In estimating the impact of this equation, a serious problem arises. That is because either $Y_1$ and $Y_2$ are not normally observed, but none of them for each household was recorded. So, the important assumption for this framework that can be stated is that individuals selected into both treatment and non-treatment groups have prospective outcome in both positions. One position is consistence while the other is not as observed. Therefore, this framework can be expressed as follows:

$$Y_i = D_i Y_{iS} + (1 - D_i) Y_{iN}$$

$$D = 1, 0 \tag{2}$$

Suppose P is the probability of observing a household with D=1, so the average treatment effect, t, can be illustrated as follows,

$$t = P \cdot \left\{ E\left(Y_S \middle| D=1\right) - E\left(Y_N \middle| D=1\right)\right\} + (1-P) \cdot \left\{ E\left(Y_S \middle| D=0\right) - E\left(Y_N \middle| D=0\right)\right\} \tag{3}$$

This equation means that the crucial problem of causal inference stems from the fact that the unobserved counterfactuals cannot be estimated (Smith and Todd, 2005). This situation requires one to employ the propensity score-matching (PSM) method in order to address this crucial problem (Rosenbaum and Rubin, 1983). Therefore, the Logistic Regression is the utmost generally used method for estimating the Propensity Score. This is used to predict the probability that an event is arisen.

$$Y\left(1 \text{ or } 0\right) = \beta_0 + \sum_i \beta_i X_i \tag{4}$$

Now when estimating the treatment effect based on propensity score, the conditional independence assumption (CIA) is required, which can be written as $(Y_N, Y_S) \perp D_i \mid X_i$, a first assumption. While in the second assumption, the average treatment of treated (ATT) is taken, which ensures that the individual with similar X values as explanatory variables have positive possibility of being a participant and non-participant (Heckman et al., 1997). Here the average treatment effect on treated can be illustrated as follows,

$$ATT = E\left(Y_{iS} - Y_{iN} \middle| D=1\right)$$
$$= E\left\{ E\left(Y_{iS} - Y_{iN} \middle| D_i=1, Y_i\left(1 \text{ or } 0\right)\right)\right\}$$
$$= E\left\{ E\left(Y_{iS} \middle| D_i=1, Y_i\left(1 \text{ or } 0\right)\right) - E\left(Y_{iN} \middle| D_i=0, Y_i\left(1 \text{ or } 0\right)\middle| D=1\right)\right\} \tag{5}$$

The first term is the treatment effect that we are going to isolate as

an average in the treatment group, which is the group of household that participated in the education of their children.

So, what will be the difference between the non-participant groups, which is the selection bias between the two groups? As the data about $E(Y_S \mid D=1)$ is already available from the participant groups, we have to find out the $E(Y_N \mid D=1)$, as the data on the non-participants support the classification of $E(Y_N \mid D=0)$ only. That is why the difference between $E(Y_S \mid D=1)$ and $(Y_N \mid D=1)$ cannot be observed for the same household members. As Rubin (1977) stated, an assumption that a set of observable covariates X, the potential outcome which is non-treatment outcomes are independent of the participation status of CIA (Conditional Independence Assumption), which is $(Y_N \perp D \mid X)$.

Therefore, after the modification of the potential outcome its mean is the same for D=1 and D=0, $(E(Y_N \mid D=1, X) = E(Y_N \mid D=0, X))$. This will allow us to use matched non-participant household members to explain how the participating group members would have performed, if they had not participated. Hence, we assumed that outcomes are conditionally mean independent of participation after conditioning on a set of observable characteristics.

Heckman et al. (1997) stated that between the outcomes of participants and non-participants their possibility will be a systematic difference due to many reasons. These differences may be due to the variety of unmeasured characteristics or outcome level of differences $(E(Y_S - Y_N \mid D=1))$. This might arise when participants and non-participants belong to different groups.

Angrist and Krueger (1999) worked on program evaluation and natural experiment approach. To estimate the effect of getting treatment randomly is not possible. The propensity scores are an alternative method for this procedure. In propensity score matching (PSM), creating pairs of the treatment and control components, with the same values related to propensity score perhaps covariates the disposal of all unmatched units (Rubin, 2001).

**Propensity score matching**

Propensity score matching is mostly used to match two groups of topics, but it can be estimated in more than two groups. Rosenbaum and Rubin (1985) for the first time stated the concept of Propensity Score Matching: the selection bias with principal emphasis on making casual extrapolations when the data set is based on non-random samples.

Also, the difference-in-difference approach using propensity score matching was developed by Heckman (1997). Becker and Ichino (2002) stated in their research work that propensity score method is a two stage method. The first stage as mentioned earlier consists of the probability model (Probit or Logit) which determines the propensity score of the household's characteristics. where $\beta_i$ is regression coefficient to be predicted and $X_i$ is an independent variable to be clarified. For propensity score matching, we applied the following equation:

$$P_{score} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots \beta_n X_n)}} \tag{6}$$

When propensity score is estimated, the appropriate matching technique is implemented. There are five main practices of Propensity Score Matching. The 1st is the Stratified Matching, the 2nd is Nearest Neighbor Matching, 3rd is the Radius Matching, 4th is Mahalanobis Metric Matching and the 5th and last one is Caliper Matching. While in the second step, each household group was matched up to predict households with the similar propensity score values. To estimate the average treatment effect, Nearest Neighbor Matching (NNM) was applied.

## Data and definition of the variables

### Brief history of the household survey data

The household survey data were launched in 1963, and named as household integrated economic survey; while in 1990, the questionnaire was revised based on national accounts to fulfill the deficiencies in data collection. Later on in 2003 to 2004, the survey was renamed as Pakistan Social Living Standard Measurement (PSLM) and the Household Integrated Economic Survey (HIES) segment was completed. The main idea of the PSLM data project was to collect the information in social and economic indicators of the households through survey.

In July 2004, the project was initiated and continued till June 2015. The basic purpose of this project was to assist the programs launched for the Millennium Development Goals to formulate poverty deterioration and other development plans, in the four provinces of Pakistan at district level. Therefore, the data help in providing information about social, income and indicators and the 18 targets which were mentioned by the UN for the implementation and fulfillment of Millennium Development Goals by the year 2015. The main indicators were based on demographic characteristics, education, health, employment, assets and water supply and sanitation (PBS, 2017).

Pakistan Bureau of Statistics had developed its own urban and rural area sampling frame in Pakistan. The cities and towns were divided into enumeration blocks, and each enumeration block consisted of 200 to 250 households. The enumeration blocks were further classified into three income groups, which were low, middle and high income, and the living standard of the people was patterned. Therefore, rural area frame work was based on a list of villages published by the Population Census Organization as part of the 1998 census.

### Getting the data and the selection of variables

The empirical analysis is based on the household level PSLM data set for 2013 to 2014. The micro- data were obtained from Pakistan Bureau of Statistics to investigate the main problems faced by households. The data consisted of four provinces of Pakistan, namely Khyber Pakhtunkhwa, Punjab, Sindh and Baluchistan, Gilgit Baldistan and Capital Territory Islamabad.

We selected the province Khyber Pakhtunkhwa from the data of all over Pakistan based on 17989 households. After removing the missing values from the data and screening the data, a total of 4388 household members were selected from Khyber Pakhtunkhwa Province for further analysis. The factors which affect the enrolment of under 20 years old household members were based on the following variables: number of household members, age of the household head, number of workers in all sectors, number of workers in the agricultural sector, highest education level of household head, distance from school, assets of the household, such as bicycle, radio, mobile phones/PCs and their literacy rate. After the selection of the variables, the data were further analyzed through Microsoft Access, Excel and R-statistics. Matching technique was applied to the model.

As the PSLM data is non-randomized and based on stratification used to investigate the factors that affect the household decision, we used the matching method. When the randomization was not possible, we did comparison between control and treatment groups, in terms of the differences in their characteristics. That is why the households who are affected by the policies or as treatment group may be different from those who are not affected (the control group).

Similarly, the treatment effect may cause outcome differences. In the field of educational research and its policies there are many covariates due to dynamic factors that affect the outcomes and statistically it is very difficult to analyze through traditional methods. As household education, number of worker, family size, distances from school, household assets and literacy rate are taken as an outcome variable; therefore to estimate such variables we investigated the balance between the treatment and control groups. Then we compared these two groups with their covariates, by using the Propensity Score Matching method as an alternative method to apply the logic of balance between the two groups.

For this purpose, we developed the model to elaborate the study, and investigated the problems. Table 1 shows the selected variables and their definitions.

## EMPIRICAL RESULTS

The empirical analysis for completion and non-completion of school in the last year involved two steps of estimation: the household members who completed school last year and those who did not. The first step consists of impact analysis tracked by a description of propensity scores for the treatment variables. To predict the probability of school completion, a logit model was introduced. The results of the propensity score matching are given in details below. Lee (2008) described that the propensity score matching is used to balance the observed dissemination of covariates between the treated and control groups.

Therefore, the success of propensity score assessment is the resultant balance. The effect of those household members who did attend school last year was further estimated through the nearest neighbor (NNM) method. The empirical results for both control and treatment group were estimated. Table 2 shows the descriptive statistics of the sample variables in each category for all the data. The treated group in Table 2 consists of 177 household members who did attend school last year with all their members and did complete school; while for control group, it was 4388 members, and consisted of household members who did not complete school last year. In Table 2, we applied the effect size based on means. When the studies for meta-analysis are based on standard deviation and means, we usually prefer standardized mean difference, raw mean difference or the response ratio for the size effect. The transformation of all effect size in standardized mean difference (d or g) is based on common metric, which thus gives us the ability to put measures of different outcome trials in the similar synthesis.

Therefore, size effect is widely used in meta-analysis as well as in primary research. The studies which are based on two arguments of the standardized mean

**Table 1.** Definitions of variables.

| Variable name | Definition | Unit |
|---|---|---|
| Y | =1 if all household members under 20 years old enrolled in school/institution and did complete the class in last year | Dummy |
| Age | Age of household head | Age |
| No of worker | Number of workers in all sectors | Number |
| No of agric work | Number of workers in the agricultural sector | Number |
| High edu | Highest education level of household members<br>Less than class 1 = 0, Class 1 to 10 = 1 to 10<br>Polytechnic diploma = 11, Associate degree = 12, Bachelor (BA) = 13<br>Post graduate (MA) = 14, Ph.D. =15 | Dummy |
| Dist to school | The closest distance from the school/institution where household member are attending | Km |
| Bicycle | =1 if household possesses a bicycle or more | Dummy |
| Radio | =1 if household possesses a radio or more | Dummy |
| PC | =1 if household possesses a PC or more | Dummy |
| Literacy rate | Rate of all family members 10 and older can read with understanding | Ratio |

Source: Pakistan social and living standards measurement survey ROUND-IX (2013-14).

**Table 2.** Descriptive statistics of sample in each category for all data.

| | Treated (Y=1) N=177 | | | | | | Control (Y=0) N=4.388 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Variable name | Mean | S.D. | Max | Min | Std mean diff | | Mean | S.D. | Max | Min | Unit |
| Hhh age | 53.08 | 13.96 | 98.00 | 18.00 | -8.16 | ** | 54.23 | 12.66 | 99.00 | 17.00 | Age |
| No worker | 2.11 | 1.47 | 8.00 | 0.00 | -29.86 | ** | 2.55 | 1.62 | 25.00 | 0.00 | Number |
| No agr work | 0.32 | 0.51 | 2.00 | 0.00 | -1.63 | ** | 0.33 | 0.55 | 2.00 | 0.00 | Number |
| Hh edu | 10.59 | 3.11 | 20.00 | 1.00 | 57.44 | * | 8.80 | 3.46 | 20.00 | 1.00 | Dummy |
| Dist. school | 1.68 | 1.50 | 7.00 | 0.00 | 29.07 | * | 1.24 | 1.32 | 7.00 | 0.00 | km |
| | | | | | | | | | | | |
| Variable name | Number of hh (=1 ) (%) | | Max | Min | Std mean diff | | Number of hh (=1 ) (%) | | Max | Min | Unit |
| Bicycle | 46.00 | ( 26.0) | 1.00 | 0.00 | 6.96 | ** | 1.006.00 | ( 22.9) | 1.00 | 0.00 | Dummy |
| Radio | 9.00 | ( 5.1) | 1.00 | 0.00 | -3.19 | *** | 254.00 | ( 5.8) | 1.00 | 0.00 | Dummy |
| PC | 160.00 | ( 90.4) | 1.00 | 0.00 | -3.43 | *** | 4.011.00 | ( 91.4) | 1.00 | 0.00 | Dummy |
| | | | | | | | | | | | |
| Variable name | Mean | S.D. | Max | Min | Mean raw diff | | Mean | S.D. | Max | Min | Unit |
| Literacy rate | 0.61 | 0.27 | 1.00 | 0.00 | *0.07* | *** | 0.54 | 0.24 | 1.00 | 0.00 | ratio |

Signif. Codes: N.S. >=0.10; * P <0.10; ** P <0.05; *** P <0.01 based on the two sample t-test.

difference are comparable (Heges and Olkin, 1985). On the other hand, the raw mean difference, which is denoted by (D) can be used as the effect size, when the scale of the outcome is either well known or characteristically evocative, because of its extensively usage. In the analysis, a part is used for the similar scale. This effect size is implemented (Borenstein et al., 2009).

The descriptive statistics of the study suggest that the standard mean difference, with its value -8.16 for the household head age has negative effect on households' school completion between the two groups. It is discussed in the previous literature that young people have the tendency to send their children to school, and they are able to complete their school. This is due to the social mobilization and the awareness by different non-profit organizations in the province that younger parents prefer to send their children to school. The reduction in the number of workers, number of agricultural worker, with its

**Table 3.** Estimated probability model in logit for PSM.

| Explanatory variables | Estimated coefficients | | p value |
|---|---|---|---|
| Intercept | 3.06E-02 | * | 0.05492 |
| Hhh age | -4.22E-04 | * | 0.06622 |
| No worker | -7.02E-03 | *** | 0.00021 |
| No agr work | 8.70E-03 | N.S. | 0.11598 |
| Hh edu | 5.75E-03 | *** | 0.00000 |
| Dist. school | 6.53E-03 | ** | 0.00286 |
| Bicycle | 1.16E-02 | * | 0.08715 |
| Radio | -6.94E-03 | N.S. | 0.56925 |
| PC | -1.70E-02 | * | 0.09786 |

Signif. codes: N.S. >=0.10; * P <0.10; ** P <0.05; *** P <0.01.

mean values is -29.86, and -1.63. This shows that, education attainment decreases the number of workers and household income. This is because, when households decide to send their children to school it decreases their work participation.

The standard mean difference for the household with its level of education has a positive effect and is increased by an average of 57.44. While the distance from school has also positive effect on completion of school, the mean difference between the treated and control group is 29.07, which shows that households can decide to send their children to long distance school.

The aforementioned discussion of this table was based on the households' demographic characteristics. The second part of the table consists of the assets of the household who possess bicycle, with its mean differences. This shows that there is positive significant relationship between completion of school and bicycle used as an asset by households. This shows that, its standard mean difference is 6.96. While owing of radio and mobile phones/PCs has negative standard mean differences ( -3.19 and -3.43) between both the treated and control groups. The last variable as literacy rate has positive relationship with school completion; its mean difference is 0.07, which is the outcome variable.

## Likelihood method and estimations

Here we estimated the propensity score matching, although it can be estimated by using models like discriminant analysis, boosted regression and probit regression (McCaffrey et al., 2004). The logistic regression is typically used for the analysis. Matching packages and Matching estimate propensity scores expended the logistic regression as the default option (Ho et al., 2011).

The fit of the model cannot be evaluated, when using the default option for estimating propensity scores. Therefore, the logistic regression is recommended to run and accomplish the model fit. Estimation of the logistic

regression in propensity score matching was recorded. The significant estimates are determined by the low p-value (that is, <0.05). Authors suggest that both statistically significant variables are related to selection (Austin et al., 2007).

Table 3 shows the empirical results of probit model. This indicates that the estimated coefficient of number of workers is -7.02, which has significant negative effect on completion of school for all the household members last year. The reduction in number of workers shows that the probability of school completion in the last year decreases the number of workers in the household. The slope of coefficient for number of agriculture workers has a positive effect, which is 8.70.

This means that the number of agriculture workers is increasing by 8.70, if the probability of school completion by a one unit is increased for all the household members. This is because when the children of the household come back from school they are engaged in agricultural work. The estimated coefficient of household education is 5.75. This shows the positive effect on the probability of school completion, and has a significant effect on household educational level.

Bjorklund and Salvanes (2011) stated in their study that there is a strong correlation between the educated households and their children. Family background strongly affects their children's education. The household takes decision about longer distance when they want to send their children to school. This is because there is a correlation between the use of bicycle and distance from school. The estimated coefficient value, 6.53 suggests that an increase occurs in distance when a unit change takes place in the probability of school completion. The use of bicycle has also positive significant effect. The estimated coefficient of the use of bicycle is 1.16 and has positive impact on school completion. This is because an increase in the number of bicycle can increase education attainment, if the schools are situated at a longer distance.

Muralidharan and Prakash (2013) by using the triple difference approach to investigate the bicycling program

**Table 4.** Descriptive statistics of sample in each category for matched data.

| | Treated (Y=1)  N=105 | | | | | | Control (Y=0)  N=201 | | | | |
| Variable | Mean | S.D. | Max | Min | Std Mean Diff | | Mean | S.D. | Max | Min | Unit |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Age | 52.78 | 11.42 | 90.00 | 34.00 | 5.91 | N.S. | 51.89 | 11.51 | 95.00 | 19.00 | Number |
| No of worker | 2.11 | 1.47 | 8.00 | 0.00 | 9.51 | N.S. | 2.18 | 1.36 | 6.00 | 0.00 | Number |
| No of agr worker | 0.32 | 0.51 | 2.00 | 0.00 | 7.42 | N.S. | 0.38 | 0.62 | 2.00 | 0.00 | Number |
| Higher edu | 10.59 | 3.11 | 20.00 | 1.00 | 6.51 | N.S. | 10.63 | 3.71 | 20.00 | 1.00 | Dummy |
| Dist to school | 1.68 | 1.50 | 7.00 | 0.00 | 3.72 | N.S. | 1.69 | 1.81 | 7.00 | 0.00 | km |
| Bicycle | 0.26 | 0.44 | 1.00 | 0.00 | 10.39 | N.S. | 0.30 | 0.46 | 1.00 | 0.00 | Dummy |
| Radio | 0.05 | 0.22 | 1.00 | 0.00 | 10.45 | N.S. | 0.06 | 0.24 | 1.00 | 0.00 | Dummy |
| Mobile/PC | 0.90 | 0.29 | 1.00 | 0.00 | 15.45 | N.S. | 0.92 | 0.27 | 1.00 | 0.00 | Dummy |

Std mean diff: N.S. >=0.10; * P <0.10; ** P <0.05; *** P <0.01 based on paired t-test, the test for equal balance in the estimated prob between treated and control (nboot = 1.000).

in rural India, stated that the treated villages, where the households benefited from the program, the schooling attainment increased by 30%. The program was launched for the girls whose schools were far from their homes. It also reduced the gender gap by 40% in secondary school level enrollment.

While in case of Mobile Phones/Pcs, negative effect on school completion and the slope of the coefficient is -1.70. This indicates that the use of mobile phones has negative effect on school completion. The probability of a unit of school completion for all members of the household reduces the use of mobile phones up to, -1.74, and has a significant effect.

## Treatment effects

Kolmogorov-Smirnov test was used for the data analysis, which was based on the bootstrap p-value. This is widely used to provide the precise estimation, even if the compared distributions are not exclusively continuous. This test provides equal balance in the estimated probability for both the treated and control groups with number of bootstraps, which is based on Monte Carlo simulations used to determine the appropriate p-value. However, we estimate the asymptotic distribution for the cases of matching estimate, when the conditional bias is ignored, and also, the matching estimators for the fixed number of matching may not extend to the semi-parametric efficiency bounds. Therefore, an asymptotic variance estimator is proposed (Abadie and Imberns, 2006). Finally, the average treatment effect of the household participating in education was assessed by comparing the deviations in individual outcomes between participants and their matched counterparts (Table 4).

Table 4 shows the descriptive statistics of sample in each category for matched data. The table is based on both treated and control groups, and their standard mean differences. The descriptive statistics for matched data show that the variables, selected for the difference

between their mean values were based on household demographic as well as assets variables. The demographic variable with their standard mean differences for the treated and control group can be illustrated as follows: the standard mean difference for the household shows that the number of workers, number of agriculture worker, households' education, distance from school, bicycle, radio, mobile phones/ Pcs has non-significant positive value. This shows that, the model is fit for the treatment effect for both the control and treated groups. Therefore, after applying the paired t-test, the test for equal balance in the estimated probability between the two groups, the outcome variable will be estimated to know the impact of literacy rate on household school completion, as shown in Table 5.

Table 5 shows the average treatment effect on both treated and control group. A total of 105 household members were in the treated group, and 201 in the control group. This shows the outcome variable, literacy rate has negative impact on household school completion. This is because school graduates have very less capabilities to read and understand a single sentence. Table 5 suggests that literacy rate for both treated and control group has a significant negative effect on school completion for all the members of the household. This is because of the poor quality of education, insufficient budget for households and distance from school to home.

Khan et al. (2016) described in their research work on the study of South Asia, that households' decision making depends on returns to education. There are other external benefits in spite of returns to education, improvement in the health status of an individual, increases the income, rise quality of life, reduces family size, increases individual productivity, political awareness, and better childcare etc. They further described that the literacy level has been increased in South Asia for the last 15 years except Pakistan, which is far behind the targeted Millennium Development Goals and has the lowest literacy rate in the region.

Mostly researchers use the returns to education as the

**Table 5.** Average treatment effects (ATT) in each matched data for literacy rate.

| | Treated (Y=1) N=105 | | | | | | Control (Y=0) N=201 | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Variable name | Mean | Max | Min | Est. Diff | AI SE | - | Mean | Max | Min | Unit |
| Literacy rate | 0.59 | 1.00 | 0.00 | -0.05 | 0.02 | *** | 0.63 | 1.00 | 0.00 | ratio |

Signif. codes: *** P <0.01 ( "AI SE" is the matching corrected standard error based on Abadie and Imbens (2005).

human productivity and a year increase in education can increase the individual earnings, while the returns and household expectation with the literacy rate as an outcome variable has not been estimated in the region. The research work shows that the low level of education attainment by the children of the household is due to lack of cognitive skills and their achievements. The students are unable to even understand a single sentence, to read and solve a simple problem. This is due to the lack of quality curriculum and untrained teachers in the region.

Therefore, the households in Khyber Pakhtunkhwa, Pakistan are not interested in the schooling of their children due to lack of quality enhancement in the education system. Household considers sending their children to school as a waste of money and time. That is why households prefer to send their children to work in some workshops and learn some skills.

## Conclusion

This work studies households' decision to participate in the education of their children. Before the analysis of the study, we were able to identify the World Bank and UNESCO data from 2016, that Pakistan has the worse situation in sending their children to school both in primary and secondary level in South Asian countries. The educational level of all South Asian countries was discussed (Khan et al., 2016).
In this study, we discussed those households who enroll their children in schools and completed their education. For this purpose, we selected Khyber Pakhtunkhwa Pakistan. We used propensity score matching method joint with the logistic regression model to estimate the situation of those households, with respect to the effect of different variables, which are: age of the household head, number of workers in all sectors, workers in agricultural sector, education of the household, distance from school, assets of the household and their literacy level.

After applying the average treatment effect on treated (ATT) of the household members, they were assessed by comparing the differences in individual outcomes between treatment and control group. It was suggested that literacy rate has a significant negative relationship with school completion. A total of 105 households from the treated group were matched with the 201 household members from control groups. This shows that the literacy rate has negative relationship with school completion and non-completion of school of the

household last year, for both control and treated groups. This is because the syllabus is not much effective according to the modern world requirement.

Low literacy has many causes, but the main reasons that effect the literacy level are as follows: socio-economic condition of the household, low household income, insufficient resources, child labor, 38% of a large number of people living below poverty line, deficiency in quality education, untrained and unqualified teachers, low level of cognitive skills, lack of facilities, and inadequate infrastructure.

However, the failure of educational policy since the existence of Pakistan is a big problem, that even educated households have lower cognitive skills. This shows the consequences of inadequate educational policies in Pakistan.

The findings suggest that without households' participation and community awareness, the decrease in the dropout from school would not be possible. Pakistan should give proper attention to their education system, which is badly affected. Poor communication and transportation system creates hurdles for people who want to send their children to school. The government should also provide well trained teachers, modern syllabus and quality environment both in public and private institutions.

## CONFLICT OF INTERESTS

The authors have not declared any conflict of interests.

## REFERENCES

Abadie A, Imbens GW (2006). Large sample properties of matching estimators for average treatment effects. Econometrica 74(1):235-267.
Agrawal T (2014). Educational inequality in rural and urban India. Int. J. Educ. Dev. 34:11-19.
Alderman H, Behrman JR, Khan S, Ross DR, Sabot R (1996). Decomposing the regional gap in cognitive skills in rural Pakistan. J. Asian Econ. 7(1):49-76.
Andrabi T, Das J, Khwaja AI, Vishwanath T, Zajonc T (2007) Learning and Educational Achievements in Punjab Schools (LEAPS): Insights to Inform the Education Policy Debate. World Bank, Washington, DC.
Angrist JD, Krueger AB (1999). Empirical strategies in labor economics. Handbook Labor Econ. 3:1277-1366.
ASER (2016). 2017,http://www.aserpakistan.org/document/learning_resources/2014/Post_2015_Agenda/MDGs%20&%EFA%20Goals.pdf.
Austin PC, Grootendorst P, Normand SLT, Anderson GM (2007).

Conditioning on the propensity score can result in biased estimation of common measures of treatment effect: a Monte Carlo study. Stat. Med. 26(4):754-768.

Barro RJ, Lee JW (2001). International data on educational attainment: updates and implications. Oxford Econ. Papers 53(3):541-563.

Becker GS (1962). Investment in human capital: A theoretical analysis. J. Political Econ. pp. 9-49.

Becker SO, Ichino A (2002). Estimation of average treatment effects based on propensity scores. Stata J. 2(4):358-377.

Bjorklund A, Salvanes KG (2011). Chapter 3-Education and Family Background: Mechanisms and Policies. volume 3 of Handbook of the Econ. Educ. pp. 201-247.

Borenstein M, Hedges LV, Higgins J, Rothstein HR (2009). *References* (pp. 409-414). John Wiley & Sons, Ltd.

Education for All (2014). National Review, Ministry of Education, Trainings and Standards in Higher Education Academy of Educational Planning and Management Islamabad, Pakistan June, 2014 (available at: http://unesdoc.unesco.org/images/0022/002297/229718E.pdf).

Creswell JW (2013). Research design: Qualitative, quantitative, and mixed methods approaches. Sage publications.

Emerson PM, Souza, A. P. (2002). Birth order, child labor and school attendance in Brazil. Nashville: Vanderbilt University, Department of Economics.

Epstein JL, Jansorn NR (2004). School, Family and Community Partnerships Link the Plan. Education Digest: Essential Readings Condensed For Quick Review, 69(6):19-23.

Gertler P, Glewwe P (1990). The willingness to pay for education in developing countries: Evidence from rural Peru. J. Public Econ. 42(3):251-275.

Graesser A, Chipman P, Leeming F, Biedenbach S (2009). Deep learning and emotion in serious games. Serious games: Mechanisms and effects pp. 81-100.

Guisinger SE, Henderson JW, Scully GW (1984). Earnings, rates of return to education and the earnings distribution in Pakistan. Econ. Educ. Rev. 3(4):257-267.

Hanushek EA, Kimko DD (2000). Schooling, labor-force quality, and the growth of nations. American economic review, 1184-1208.

Haque NU (1977). An economic analysis of personal earnings in Rawalpindi City. Pakistan Dev. Rev. 16(4):353-382.

Heckman JJ, Ichimura H, Todd PE (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. Rev. Econ. Stud. 64(4):605-654.

Ho DE, Imai K, King G, Stuart EA (2011). Match It: nonparametric preprocessing for parametric causal inference [Internet]. *J. Stat. Soft.*

Horowitz AW, Souza AP (2011). The impact of parental income on the intra-household distribution of school attainment: A measurement strategy and evidence. Q. Rev. Econ. Financ. 51(1):1-18.

Hallak J (1995). Educational Planning: The International Dimension (Vol. 1032). Taylor & Francis.

Kadzamira E, Rose P (2003). Can free primary education meet the needs of the poor?: evidence from Malawi. Int. J. Educ. Dev. 23(5):501-516.

Karemesi J (2010). Universalising Primary Education in Kenya, Is it Sustainable. *Surfing the Web: http://usbec. wordpress. com.*

Khan S, Puangkaew L, Shusuke M (2016). The contribution of international community and the Dynamics of History in Educational Policies in Pakistan: the case of success or failure, Int. J. Curr. Res. 8(2):42150-42156.

Lee DS (2008). Randomized experiments from non-random selection in US House elections. J. Econometrics 142(2):675-697.

Lloyd CB, Mete C, Grant MJ (2009). The implications of changing educational and family circumstances for children's grade progression in rural Pakistan: 1997–2004. Econ. Educ. Rev. 28(1):152-160.

Mansory A (2007). A Study of the Drop Out Rate in Afghan Schools. Kabul: Swedish Committee for Afghanistan (SCA).

McCaffrey D, Ridgeway G, Morral A (2004). Propensity score estimation with boosted regression for evaluating adolescent substance abuse treatment. Psychol. Methods 9(4):403-425.

Mincer J, Polacheck S (1974). Family investments in human capital: Earnings of women. In Economics of the Family: Marriage, Children, and Human Capital (pp. 397-431). University of Chicago Press.

Muralidharan K, Prakash N (2013). Cycling to school: increasing secondary school enrollment for girls in India (No. w19305). National Bureau of Economic Research.

Munda SW, Odebero S (2014). The influence of education costs on students' academic performance in Kenya: An empirical study of Bungoma County Secondary Schools. Asian J. Educ. Res. 2(1).

Muralidharan K, Prakash N (2013). Cycling to school: increasing secondary school enrollment for girls in India (No. w19305). National Bureau of Economic Research.

Odaga A, Heneveld W (1995). Girls and schools in Sub-Saharan Africa: From analysis to action (Vol. 298). World Bank Publications.

Pakistan Bureau of Statistics (2017) http://www.pbs.gov.pk/content/pakistan-social-and-living-standards-measurement

Patton MQ (1990). Qualitative evaluation and research methods. SAGE Publications, inc.

Patrinos HA, Psacharopoulos G (2002). Returns to investment in education: a further update. World Bank Policy Research Working Paper Series.

Pritchett L, Beatty A (2015). Slow down, you're going too fast: Matching curricula to student skill levels. Int. J. Educ. Devel. 40:276-288.

Rosenbaum PR, Rubin DB (1983). The central role of the propensity score in observational studies for causal effects. Biometrika 70(1):41-55.

Rosenbaum PR, Rubin DB (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. Am. Stat. 39(1):33-38.

Rubin DB (1977). Assignment to Treatment Group on the Basis of a Covariate. J. Educ. Behav. Stat. 2(1):1-26.

Rubin DB (2001). Using propensity scores to help design observational studies: application to the tobacco litigation. Health Services and Outcomes Research Methodology, 2(3-4):169-188.

Sathar ZA, Lloyd CB (1994). Who gets primary schooling in Pakistan: Inequalities among and within families. The Pakistan Development Review, 103-134.

Shah MAA, Anwar M (2014). Impact of Parent's Occupation and Family Income on Children's Performance. Int. J. Res. 1(9):606-612.

Thomas K, Perry-Jenkins M (1994). All in a day's work: Job experiences, self-esteem, and fathering in working-class families. Family Relations, pp. 174-181.

UNESCO (2012) Pakistan Education for All review report 2015; 2014, http://unesdoc.unesco.org/images/0022/002297/229718E.pdf

UNESCO (2016). Education Statistics all Indicators. http://data.uis.unesco.org/

World Bank (2016). Education Statistics all Indicators,http://databank.worldbank.org/data/reports.aspx?source=Education-Statistics-~-All-Indicators

Zeira A, Dekel R (2005). The self-image of adolescents and its relationship to their perceptions of the future. Int. Soc. Work, 48(2):177-191.