Why Does a Child Leave School? Analyzing Dropout Reasons Using Household Data

Haris Sohail & Waleed Effendi

Mushtaq Ahmad Gurmani School of Humanities and Social Sciences

Lahore University of Management Sciences (LUMS)

Abstract

The following paper addresses the daunting issue of children dropping out of school without completing their education. We use household data from the Pakistan Social and Living Standards Measurement (PSLM) survey and use linear probability model (LPM) and logistic regression (logit) model estimations to analyze the factors affecting the decision of a child to drop out from school. Descriptive statistics indicate that the average child who drops out does so at 14 years of age, which is also the minimum age an employee has to be. Building a model with control variables added using previous relevant literature, our results show that the probability of a child dropping out of school increases with age and decreases as the economic situation of his/her household improves. A boy is less likely to drop out than a girl holding all else constant, while a rural child is more likely to drop out than an urban child, ceteris paribus.

1. Introduction

Education stands as a fundamental pillar in the development and progress of any society (Wu & Wu, 2019). It helps boost the productivity of a nation's workforce, improves their entrepreneurial ability, and equips them with the knowledge and skills necessary for personal growth. When the labor force of a nation invests in itself and is in possession of the skillset necessary for advancement, the consequence is the economic prosperity of a nation. Understanding its crucial importance in boosting economic growth, enhancing social mobility, and alleviating poverty, basic education was affirmed as a fundamental human right by the Universal Declaration of Human Rights.

The first comprehensive framework for understanding the importance of education from an economic perspective was offered by the Nobel-winning economist Gary Becker in the 60s. His analysis of acquiring education was similar to investment in physical capital, like machinery or

equipment. According to Becker (1964), individuals invest in themselves by spending on their education which makes them more productive and efficient, and the return on this investment is in the form of higher earnings over the lifetime of individuals. Although Becker had researched related topics such as the application of economic analysis to human areas such as discrimination, marriage, etc., his research on human capital was considered by the Nobel committee to be his most noteworthy contribution to economics (source: NBER).

Becker's Human Capital Theory – as it came to be known – highlights the importance of investing in human capital as a driver of individual and economic growth. Despite all the dividends that education brings, a sizeable part of humanity is deprived of even the elementary education levels (Amir-ud-Din et al., 2022). Moreover, a significant proportion of students enrolled in educational institutions is likely to drop out of school with far-reaching consequences (ibid).

In Pakistan, the situation is extremely alarming and was brought to the spotlight by a UNICEF report published in 2020. The gravity of the situation can be understood from the fact that 22.8 million children in Pakistan aged 5–16 years (almost 44% of the total children of school-going age) remained out of school in 2018 making Pakistan the country with the second largest out of school children's population in the world (Unicef, 2020). Another point highlighted was that almost 20% of the students dropped out before even completing primary grades.

In this paper, we attempt to analyze the factors behind a child dropping out of school. The paper began with an introduction to the problem of dropouts – and the seriousness of it in the context of Pakistan - and how they hamper the economic prosperity of a nation owing to education's role in fostering growth. Next, we analyze the relevant literature which will be followed by regression analysis, discussion of the results and the policy implications they entail.

2. Literature Review

In Pakistan, as in many developing countries, education stands out as a crucial instrument for driving economic growth, alleviating poverty, and fostering social inclusion. The country has made significant efforts to enhance access to education, improve the literacy rate, and close the huge gender gap that exists even at primary and secondary levels.

In collaboration with the World Bank, the country launched a conditional cash transfer scheme for lower secondary girls in rural areas of Punjab in 2003, which significantly increase the share of female enrolment in the targeted zones. This program was one of many cash transfer initiatives launched in partnership with international donor agencies. Another example is the Education Voucher Scheme launched by the Punjab Education Foundation in 2006 that provides free education to eligible students from slum areas at primary and secondary levels.

The performance of these programs was evaluated by a UNICEF report published in 2013 as part of a global initiative on Out-of-School Children (OOSC) in 26 countries including Pakistan launched through a joint effort by UNICEF and the UNESCO Institute for Statistics (UIS). The goal of the initiative is to improve statistical information and analysis of out-of-school children (OOSC) and to scrutinize the factors leading to exclusion from schooling and the policies related to enhancing participation. According to the report, these and other efforts by the Pakistani

Government and NGOs have enhanced access to education - children are more likely to go to school today than in previous generations— however, education performance across the country remains marked by deep disparities based on gender, geographic location and wealth (Unicef, 2013). Two types of constraints have been identified by the report for this abysmal performance:

Demand-side barriers:

Demand-side socio-cultural barriers and bottleneck influencing exclusion from school in Pakistan are strongly related to society's attitude to gender roles, with girls in particular facing restrictions on their mobility because of fears about their safety and the need for them to carry out household duties, and boys experiencing pressure to start contributing to household incomes. Girls also are also vulnerable to early marriage and sexual harassment, causing them to drop out of school. Lack of awareness of the importance of starting and continuing education also has a profound influence, especially at pre-primary level. Poor health can also prevent children from obtaining a full cycle of education.

Demand-side economic barriers center on family poverty. The costs of schooling include not only direct costs such as expenditure on school materials, examination fees and transportation but also indirect costs related to the opportunity cost of a child's time. Boys are often compelled to drop out of school in order to work and girls are required to help in the household. Child labor is common for children aged 10–14 years; the vast majority of it is unpaid.

Supply-side barriers:

Supply-side barriers and bottlenecks that result in children being excluded from school or in dropping out early relate to the undersupply of schools, particularly at pre-primary and lower secondary levels; inadequate school infrastructure and facilities, including a lack of buildings, water supply, toilets for boys and girls, boundary walls and electricity; problems with teacher supply, deployment and training; irrelevant and out-of-date curriculum and textbooks; deficiencies in the teaching–learning process; and harsh corporal punishment. These barriers and bottlenecks are more serious in rural areas than urban areas and in girls' schools than boys' schools, especially at lower secondary level.

Political, governance, capacity and financial barriers and bottlenecks affect the smooth functioning of the education system. The most important barriers of this type affecting the exclusion of children from education in Pakistan are problems related to School Management Committees (SMCs); issues with devolution including unclear roles and responsibilities between provincial and district authorities; weak coordination and implementation of programs; problems with monitoring and data collection; lack of clarity and non-regulation of non-public provision of education; lack of political commitment; and problems with budgetary allocations including under-spending, overspending, unpredictability and inconsistency.

The influence of child labor on dropout rates.

Even though child labor is part of demand-side constraints to the attainment of education, child labor and the involvement of children in economic activities and household chores is an important component in the analysis of OOSC, since prior research has demonstrated that many children who are not in school are engaged in some form of economic activity. In fact, child labor, which is

strongly linked to the socioeconomic background of households, may be one of the primary reasons that keeps children out of school or causes them to drop out. The International labor Organization's Convention No. 138 stipulates a minimum age of generally 15 years for admission to employment or work (Article 2).

Empirical works.

The first empirical work we analyze is the 2013 UNICEF report. It uses multinomial regressions to isolate the impact of individual and household control variables on schooling and child labor outcomes. The controls include household head's education, household head's sector of employment, household head's gender, mother's education, mother's employment status, household size, child's age and gender, type of family (nuclear or joint), etc. The data was obtained from eight rounds of the Household Integrated Economic Survey (HIES) i.e., 1990-91, 1992-93, 1993-94, 1996-97, 1998-99, 2001-02, 2005-06, and 2007-08. The household survey data was supplemented by Labor Force Surveys (LFS) data for the corresponding years. The main findings of the report included:

- Sons of literature mothers are 7.3 times less likely to become child laborers than the sons of illiterate mothers,
- As the household head's education level increases, the probability of a child going to work declines,
- The probability of child labor significantly increases with age; an 11-year-old is nearly 1.5 times more likely to go to work than a 10- year-old,
- The most important result – in the words of the author – was the differential impact of probabilities of child labor for boys and girls: holding all else constant, girls are nearly 10 times less likely to go to work than boys.

Another empirical work is an analysis of the risk factors of school dropouts in Pakistan by Amir-ud-Din et al. (2022). They use data from the Pakistan Demographic and Health Survey (PDHS) 2017-18 and employ multivariate logistic and multinomial logistic regression models. Their results suggest students are more likely to drop out because of a lack of interest when they were enrolled in the lower educational levels, were orphaned, lived in the urban areas, were married, belonged to households where the household head was above 40 years or was female.

A related study is attributed to Satti & Jamil (2021) who used data from the Pakistan Social and Living Standards Measurement (PSLM) survey for the year 2014-15. The sample used was composed of 119,101 children and the results show that a female child is more likely to drop out than a male child, an educated or employed mother reduces the probability of dropping out. Other factors affecting the decision to dropout were identified as the economic situation of the community the child belongs to, the gender of the household head, household wealth, per capita household income, and distance from middle or secondary schools.

3. Data and Methodology

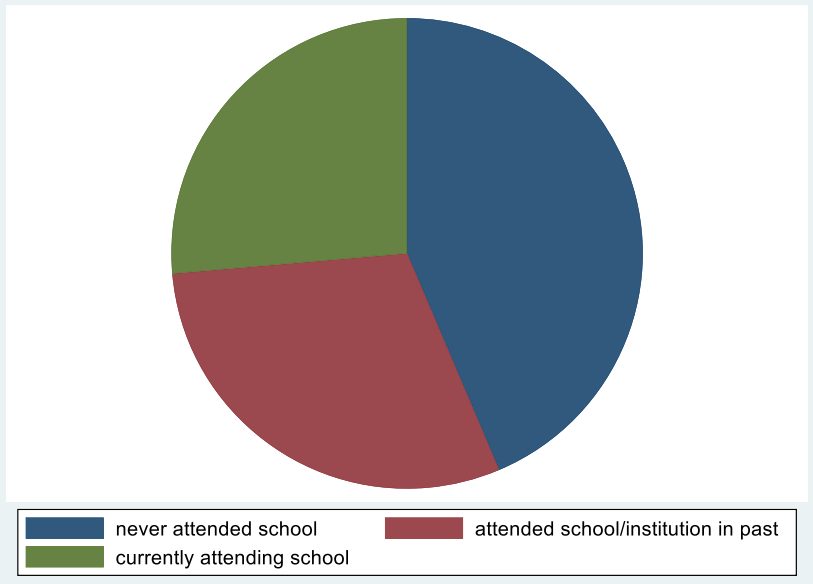
The dataset used for our analysis is the Pakistan Social and Living Standards Measurement (PSLM) survey 2019-20 and the associated Household Integrated Economic Survey (HIES). The PSLM District Level Survey is the flagship project of the Pakistan Bureau of Statistics (PBS) and is designed to provide data for the monitoring of 21 SDG (sustainable development goals) indicators. This dataset will allow us to develop key insights regarding dropouts and help draw conclusions to formulate policies from. The regression analysis includes the LPM (Linear Probability Model) and Logistic Regression Model which we delve deeper into later in the next section.

4. Results

The following section begins with descriptive statistics and is followed by the estimation of dropout rates using the qualitative dependent variable models, which include the linear probability model (LPM) and the logistic regression model. Overall, the dataset is composed of more than 160,000 households and a total of 876,355 individuals. The manual of instructions distinguishes between single-person households and multi-person households, the distinction evident from the names.

Our analysis begins with the third section of the questionnaire which is related to the education of the respondent. The second sub-section is related to formal education (as opposed to vocational training) and begins with a question about the respondent's educational background. A total of 814,276 responses were recorded and categorized to either of the following three groups: never attended school, attended school in the past, or currently attending school. The breakdown of the answers is provided in the pie chart below. The majority of the respondents – a massive 44% - had never attended school.

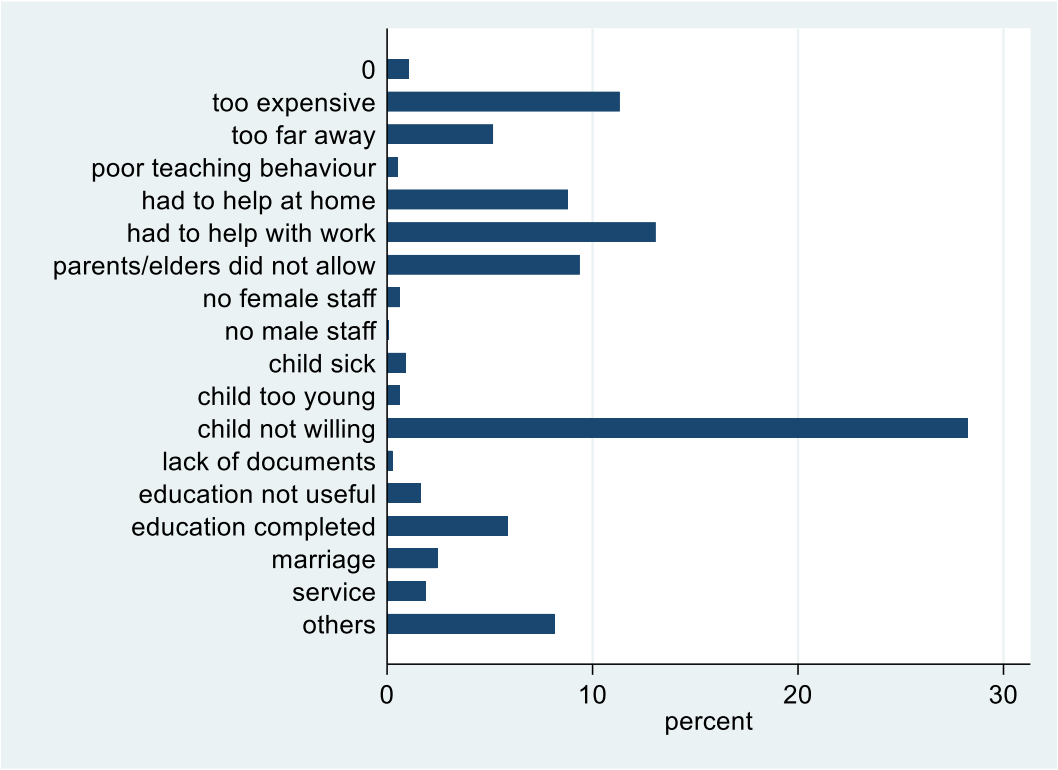
Figure 1. Composition of individuals in the dataset based on enrollment status.



Source: Authors' estimation from the data provided by PSLM 2019-20.

We are interested in two of the categories mentioned: those who are currently attending school and those who attended a school or an institution in the past, i.e., those who have dropped out. Those who dropped out were asked to explain the reasoning behind leaving the school and the responses were assigned the most appropriate code based on ten different categories. These categories and the proportion of individuals whose reasons were coded under those specific categories are given in the bar chart below. As evident from the graphic, the three most cited reasons for dropping out are unwillingness of the child, requirement of help with work, or school being unaffordable.

Figure 2. Reasons for dropping out.



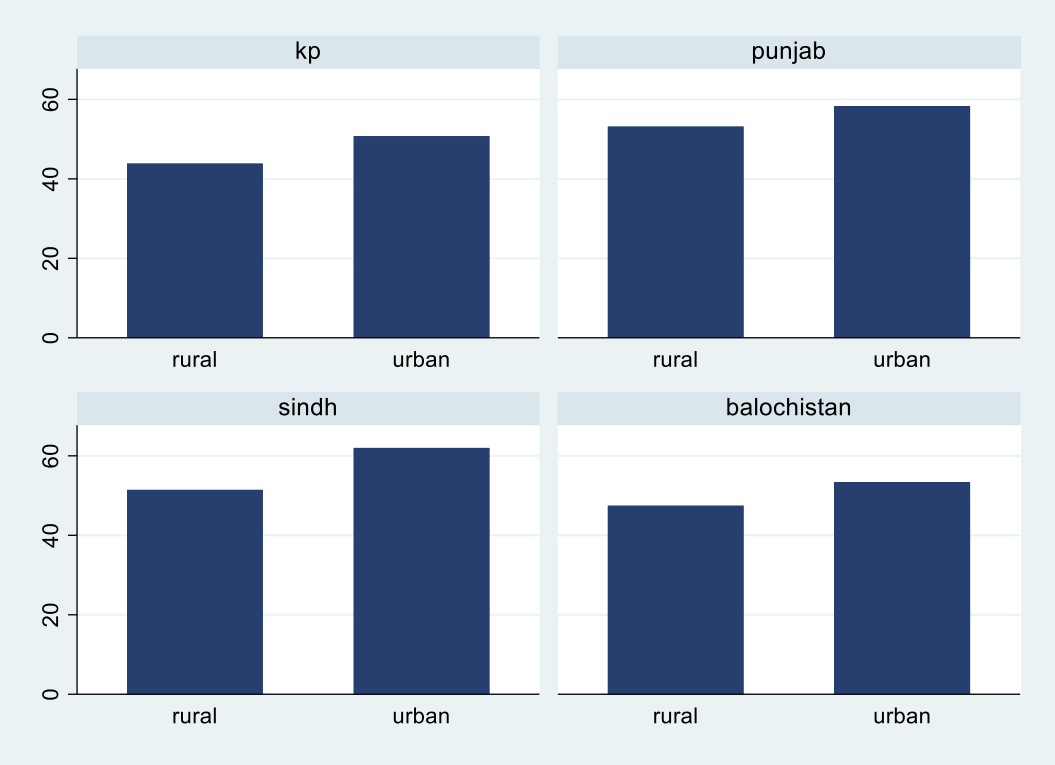
Source: Authors' estimation from the data provided by PSLM 2019-20.

Based on the reasoning of the individuals, it seems intuitive to infer that one of the main determinants of the individual's decision to drop out will be the household income. The problem is that the survey does not report household income as a separate variable. Although we have tried to construct the required variable – by summing up the individual's income from different sources such as paid labor, agricultural income, rental income, benefits from pension, etc., and then adding up the income of all the individuals in the household to construct household income – the merging process of the income and education datasets results in the loss of tens of thousands of observations. Therefore, the variable we use in our model is the household's economic situation which is categorized as "Same as Before", "Better than Before", or "Much Better than Before".

Next, we drop the individuals who had never attended school. This has been done by Satti & Jamil (2021) and the reasoning behind it is that we are interested solely in why individuals leave school

once they start it. This has left us with almost 460,000 observations out of which 53% left school at some point. We constructed a dummy variable for dropping out of school to use later in the analysis. The graph below shows the dropout rates by provinces and regions (urban and rural). Dropout rates everywhere are between forty and sixty percent of the total sample who has been to school, with the exception of Sindh Urban where dropout rates are higher than 60%.

Figure 3. Dropout rates by provinces split by regions.



Source: Authors' estimation from the data provided by PSLM 2019-20.

To begin regression analysis, the first step is the generation of the control variables. The PSLM data files are labeled according to the corresponding sections in the questionnaire. We generate the region dummy and merge the education data file with the household roster data file. Since we are interested in why 'children' drop out, we remove all individuals older than sixteen years of age. This leaves us with almost 200,000 observations and a mean dropout rate of almost 8%. The average age at which a child drops out is 14 - the minimum age a child has to be to get employed mandated by the Constitution of Pakistan (Article 37) – and this means that the average child who drops out of school does it as soon as she/he is old enough to work.

Moving to gender, 57% of individuals in our remaining sample are males and the female dropout rate is higher than the male dropout rate. We generate a dummy for gender that equals one if the child is a boy and zero otherwise. Similarly, we generate a region dummy that equals one if the household is in the urban region. Similar to Amir-ud-Din et al. (2022) and Satti & Jamil (2021), we include a dummy for the gender of the household head, it will equal one if the household head is

male and zero otherwise. Next, we merge the dataset with section 'g' that include information on the assets and financial condition of the household. This section has household-level data so we change the correspondence in the merging process and obtain information on the self-assessed economic situation of the household. Finally, this is the regression equation we will be estimating:

$$Dropout_{ih} = \beta_0 + \beta_1 Region_h + \beta_2 Gender_i + \beta_3 Age_i + \beta_4 Economic\ Situation\ of\ the\ Household_h + \beta_5 Gender\ of\ the\ Household\ Head_h + \mu_{ih}$$

Where the decision to drop out of child *i* of household *h* depends on the region in the household is located in, the age of the child, the gender of the child and the head of the household the child belongs to, and the economic situation of the household the child belongs to. Here drop out is defined as a binary variable that equals one if the child drops out. The linear probability model estimation of the equation yields the estimates in the table on the next page.

The LPM is estimated on more than two hundred thousand observations and reports an R-squared of 0.111. The second model postulates that a girl is 2.42% more likely to drop out than a boy, all else constant, highlighting the gender disparity that exists in Pakistan's education system. Children growing in an urban household are almost 2.5% less likely to drop out of school than children growing up in rural households, ceteris paribus. With age, the probability of leaving school increases significantly; more specifically, as age increases by a year, the probability of dropping out increases by almost 2.6%. The probability of dropping out decreases by 1% as the self-assessed economic situation increases by one point (the scale has been defined previously). The gender of the household head does not impact the decision-making process behind the child's dropout decision as reflected by the insignificance of its coefficient. Balochistan, which was our base category, suffers the worst in respect of educational outcomes because a child residing in KPK is 3% while a child residing in Punjab is 3% less likely to drop out than a child residing in Balochistan.

The LPM has the advantage of ease of interpretation. However, it is not the econometrician's recommended tool because firstly, it predicts probability values greater than one or less than zero, and secondly, it assumes a constant slope. Assuming a constant slope translates into the fact that it does not take into consideration the point that below for example, 6 years of age, the probability of dropping out decreases only marginally. To address this problem, we estimate the logistic regression model - or the logit model as it is widely known – which is based on the logistic distribution and yields a shape similar to the cumulative distribution function.

Interpretation of the logit model estimates becomes complex because the dependent variable in the logit model is the natural log of the odds ratio of the probability of the event. In this case, the event would be the child dropping out, so the dependent variable is the natural log of the odds of the child dropping out.

Table 1. LPM estimates (with the addition of provincial dummies).

VARIABLES	(1) dropout	(2) dropout
Region (urban)	-0.0186***	-0.0248***
	(0.00122)	(0.00127)
Gender (male)	-0.0239***	-0.0242***
	(0.00113)	(0.00114)
Age	0.0257***	0.0258***
	(0.000164)	(0.000164)
Economic situation	-0.00906***	-0.00866***
	(0.000574)	(0.000574)
Male Head	0.00166	0.00219
	(0.00201)	(0.00202)
KPK		-0.0305***
		(0.00232)
Punjab		-0.0111***
		(0.00214)
Sindh		0.00406
		(0.00247)
Constant	-0.141***	-0.129***
	(0.00304)	(0.00360)
Observations	202,998	202,998
R-squared	0.111	0.112

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The coefficients estimate using the logit model estimation are given in the table below. Note that since logit uses log odds ratio as the dependent variable, we need to convert the estimates into odds ratio by taking the exponent.

Table 2. Logit estimates and Odds Ratio's.

VARIABLES	(1) Logit	(2) Odds ratio
Region (urban)	-0.368***	0.692
	(0.0204)	
Gender (male)	-0.372***	0.6895
	(0.0181)	
Age	0.489***	1.63
	(0.00396)	
Economic situation	-0.165***	0.845
	(0.00933)	
Male Head	0.0416	1.042
	(0.0337)	
Constant	-7.727***	
	(0.0667)	
Observations	202,998	202,998

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The interpretation of the odds ratio is such that if the odds ratio is greater than one, then the variable increases the odds and thus the probability of the event. Both age and a male household head increase the probability of the child dropping out although the coefficient for the gender of the household head is insignificant. Similar to the previous estimations, the better the economic situation of a household, the more likely a child it is that the child stays in school.

5. Conclusions and Policy Implications

The policy implication our study entails is that the Pakistani government needs to focus more on reducing the prevalent gender gap in the education sector. The disparity between male and female literacy evident from our analysis must be dealt with immediately else it will become a daunting issue later on. The government must continue with its conditional cash-transfer schemes and other initiatives focused on improving female enrolment, especially in the rural areas of the country.

The divide between the urban and rural areas of Pakistan with regard to the availability of schools, qualified teachers, and adequate facilities and infrastructure is also evident from the analysis. The

government stands to gain a lot from diverting its attention to rural support programs and institutes of rural development which it has ignored for long.

The significance of provincial dummies points to the differential treatment the provinces receive when it comes to developmental expenditure. This must be addressed instantly else there is a risk of this discrepancy being perceived as discrimination by the provinces and could ignite ethnic tensions.

Age was identified as the most significant determinant of whether or not a child stays at school. A child leaving school as soon as he or she is old enough to work means families do not believe continuing education will be worthwhile for the child i.e., the return on investment is not thought to be positive. To help attenuate the situation, the government must focus on improving the job market, spreading awareness about the marginal rate of return on education, and how good schooling is the key to social mobility and the bridge to a better life.

References

Amir-ud-Din, R., Mahmood, H. Z., Abbas, F., Salman, V., & Zafar, S. (2022). Leaving studies because of lack of interest: An analysis of the risk factors of school dropouts in Pakistan. *Quality & Quantity*, *56*(5), 3189–3214. https://doi.org/10.1007/s11135-021-01266-9

Becker, G. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. University of Chicago Press.

Education: Giving every child the right to education. (2020). Unicef. https://www.unicef.org/ pakistan/education

ILO CONVENTION No. 138 AT A GLANCE. (n.d.). International Labour Organization. https://www.ilo.org/sites/default/files/2024-04/C138_at_a_glance_EN.pdf

Out-of-School Children in the Balochistan, Khyber Pakhtunkhwa, Punjab and Sindh Provinces of Pakistan (Global Initiative on Out-of-School Children). (2013). Unicef.

Satti, R., & Jamil, M. (2021). Socio-Economic Determinants of School Dropouts: An Evidence from Households in Pakistan. *IRASD Journal of Economics*, *3*(3), 388–401. https://doi.org/10.52131/joe.2021.0303.0053

Wu, E., & Wu, R. (2019). *Collaborative Model of Emerging Technologies in Asia Pacific*.