Statistical Modelling of Education, Training and Labour Market Outcomes Among Working-Age Indians

Samapriya Trivedi*[1], Amitabh Wahi[2]

[1*]Assistant Professor, Department of Statistics, Email- s.p.trivedi796@gmail.com

[2]Professor, Department of Physics, Amity School of Applied Sciences, Amity University Uttar Pradesh, Lucknow Campus, Lucknow, Uttar Pradesh, India

HOW TO CITE:

Samapriya Trivedi, Amitabh Wahi (2026). Statistical Modelling of Education, Training and Labour Market Outcomes Among Working-Age Indians International Journal of Special Education, 41(2), 766 - 786

ABSTRACT:

Education and skill formation are central to labour market inclusion, yet the relationship between schooling, training and employment outcomes in India remains uneven across social and economic groups. The objective was to examine how education level, technical education, and vocational or technical training are associated with labour force participation, work participation, and employment-type-specific earnings among working-age individuals in India. PLFS 2023–24 unit-level data for individuals aged 15–59 were analysed using survey probability weights. Descriptive statistics, chi-square tests, Cramér's V and four nested survey-design logistic regression models were applied. Regression analysis was conducted only for labour force participation, with models progressively adjusting for education/training, demographic factors, social controls, and State/UT fixed effects. The results showed substantial variation in labour force participation. Male LFPR was 83.52%, compared with 45.20% for females. Rural LFPR was 67.59%, compared with 57.02% in urban areas. Education showed a non-linear participation pattern: diploma/certificate holders had the highest LFPR at 84.33%, while higher secondary respondents had the lowest at 48.52%. Vocational/technical training showed the strongest association with LFPR, with Cramer's V = 0.421. In the fully adjusted logistic model, technical education remained positively associated with LFPR, while vocational/technical training showed the strongest adjusted association. Earnings patterns differed by worker type: education was strongly rewarded in regular wage/salary work, moderately in self-employment, and weakly in casual daily wage work. The findings suggest that education alone is insufficient for labour market inclusion. Technical and vocational pathways, gender-sensitive employment policy, and employment-type-specific labour strategies improve India's education-to-work transitions and provide a framework for other developing economies.

Keywords: Vocational training, labour force participation, Logistic regression, PLFS, India

Introduction

Education and training are critical to all forms of development. Their effects are evident in the duration of schooling and the transition from learning to working. India's transition remains lopsided. The expansion of formal schooling has not been translated into participatory employment, decent work, and equitable earnings. These are differentiated by levels of education, gender, locality (rural versus urban), social and religious affiliations, and access to technical and vocational training. India's education policy research focuses on educational attainment and its value to the labour market. It is critical to ascertain whether formal schooling incorporates cognate skills, whether school leavers can access the labour market, and whether various forms of work provide decent, occupationally congruent employment. Earlier studies on India show that education, vocational training and labour market outcomes are closely connected, but the relationship is not always linear or uniform across population groups and employment types (Agrawal, 2013; Agrawal & Agrawal, 2017; Ahmed, 2016; Kumar et al., 2019).

The issue is especially relevant to inclusive and special education research because inclusive education should be understood as more than access to classrooms. A stronger educational system must also support social participation, employability, skill development, and fair transitions into adulthood. In a highly segmented labour market such as India's, general education, technical education and vocational training may operate differently for regular employees, casual workers and self-employed workers. Therefore, a study that separates labour force participation, the worker-to-population ratio, regular wages or salaries, casual daily wages, and gross self-employment earnings can provide a more precise account of how education and training are associated with labour market outcomes. This is also consistent with the proposed workflow, which treats participation and the three income categories as separate outcomes rather than combining them into a single wage measure.

The existing literature provides several reasons for studying these relationships statistically. Human capital theory suggests that education raises productivity and earnings, and global evidence continues to show positive returns to education (Psacharopoulos & Patrinos, 2018). However, Indian evidence is more complex. Returns to education vary by gender, caste, religion, region and employment status (Rani, 2014; Mitra, 2019; Hussain & Mukhopadhyay, 2023). Recent PLFS-based work also highlights the need to distinguish among regular, casual, and self-employed labour markets, as education may yield different returns across these segments (Chen et al., 2022; Aakanksha & Bishnoi, 2024). Vocational training impacts self-employment and female labour force participation (Bairagya et al., 2021; Choudhury et al., 2024; Refeque & Azad, 2022). Mishra and Trivedi (2024) found that India's urban gender wage gap exists across all quantiles of the wage distribution and can be examined using distribution-sensitive statistics.

Another issue is that comparisons of a descriptive nature are inadequate for disentangling the education-training nexus from other related dimensions. For example, education level may differ systematically by sex, social group, religion, sector, and State or Union Territory. Similarly, those who receive vocational or technical training may differ from those who do not in age, marital status, location, and background. Without control for these other factors, the differences in labor market involvement or income may be due to these other factors, and not education or training. This points to the need for a sophisticated survey design and related statistical methods. A chi-square test may be used along with Cramer's V to describe the relationship between categorical variables. Logistic regression may be used to estimate the relationship between education, training, and labor market participation. Due to the nature of the labor market, a separate model is needed to represent income.

The gender factor is of critical importance here. It has been shown that several factors explain the participation of the female labor force in India, namely, education, the structure of the household, marriage, widowhood, social norms, care work, and labor market demand (Chatterjee et al., 2018; Chatterjee & Vanneman, 2022; Desai &

Joshi, 2019; Reed, 2020). It is clear from these studies that the expansion of education has not led to a parallel increase in labor market participation. Education, social norms, and labor market demand are intertwined. Therefore, in a statistical model that captures education and employment in India, gender and marital status must be considered as the principal demographic variables.

Analytical skills are also required to study vocational and technical training. Some studies on Indian literature state that vocational education and training can improve one's employability, but training access is inconsistent, and its labor market outcomes are inconsistent depending on the type of program, the domain, and the social class of an individual (Agrawal, 2013; Ahmed, 2016; Kumar et al., 2019; Schneider, 2024). From an international perspective, systematic evidence on technical and vocational education and training suggests that training interventions may affect employment outcomes. However, outcomes are highly variable depending on the contexts and designs of the training programs (Tripney & Hombrados, 2013). For this reason, it is essential to ask whether vocational and technical education and training are valuable in contexts where demographic and social variables are controlled.

With the above in mind, this research intends to give a statistically oriented study of education, training, and labor market outcomes in India using the 2023-24 PLFS unit record data. The first goal is to identify labor force and work participation differentials by education level. The second goal is to examine the impact of vocational or technical training on labor force participation. The third goal is to analyze the differences in earnings by education level among regular, casual, and self-employed individuals. The fourth goal is to run logistic regression to model labor force participation using education, technical education, vocational training, gender, age, marital status, social category, religion, sector, and State/Union Territory as covariates. The final goal is to determine whether education and vocational training have a statistically significant relationship with labor force participation.

The study uses nationally representative survey data and separates income and participation outcomes to add to the scholarship in three ways. The study uses the most up-to-date round of the PLFS. It uses a better methodology by not treating every income outcome as a generalized wage variable. Finally, it places education and training in the context of measurable routes to labour market participation, which pertains directly to inclusive education policy and research on post-school transition.

2. Methodology

2.1 Data and Variables

The unit-level data used in this study were from the Periodic Labour Force Survey 2023-2024. The Periodic Labour Force Survey was initiated by the National Sample Survey Office and aims to provide sets of employment and unemployment indicators, such as the rate of labour force participation, the worker-population ratio, and the unemployment rate. The annual round of the PLFS for the year 2023-2024 covered the period from July 2023 to June 2024 and was designed as a participant survey of a representative sample of the nation (Stats and Program Implementation). The analysis was restricted to the working-age population aged 15–59 years because this age band is most relevant to the education-to-work transition, labour force participation, and wage or earnings outcomes.

The principal outcome variable for regression analysis was labour force participation. Labour force participation was coded as a binary variable:

$$\mathrm{LFP_i} = \begin{cases} 1, & \text{if individual i was in the labour force} \\ 0, & \text{if individual i was not in the labour force} \end{cases} \tag{1}$$

The labour force included persons who were employed or unemployed but available for, or seeking, work under the relevant PLFS activity-status classification. The worker population ratio was used as a secondary participation outcome in descriptive and association analyses. It was coded as:

$$\mathrm{WPR_i} = \begin{cases} 1, & \text{if individual i was employed} \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

The main explanatory variables were education level, technical education, and vocational or technical training. Education level was entered as a categorical variable, with not literate as the reference group. Technical education was coded as a binary or categorical indicator depending on the PLFS classification, with no technical education as the reference category. Vocational or technical training was measured using the PLFS training variable, with no vocational or technical training as the reference category.

Control variables included sex, age, age squared, marital status, social group, religion, rural–urban sector, and State or Union Territory. Sex was coded with male as the reference category. Marital status was treated categorically, with never married as the reference group. The social group was treated categorically, with the Scheduled Tribe (ST) as the reference group. Religion was included, with Hindu as the reference category, and Muslim, Christian, and other religions as comparison categories. Sector was coded with rural as the reference category. State or Union Territory was included as a fixed-effect control in the fully adjusted model.

Income-related variables were analysed only descriptively in this section of the study design and were separated by employment type. Regular employee wages or salaries, casual daily wages, and gross self-employment earnings were treated as distinct income concepts. They were not combined into a single wage variable because regular employment, casual labour and self-employment represent different labour market arrangements and income-generation mechanisms.

2.2 Statistical Methods

All analyses were conducted using Stata version 13 on Windows 11 x64 bit. The PLFS sample design was accounted for using survey weights. The survey weight variable was specified as a probability weight, and all descriptive statistics, cross-tabulations, and regression models were estimated using weighted procedures. Survey weighting was necessary because PLFS is based on a complex probability sample rather than a simple random sample. The unequal selection of sampling units was accounted for in the sample through weighted estimation, which made the sample observations more representative of the target population. In complex surveys, appropriate survey weights and a design-based approach to estimation were required. Unweighted data, especially in the context of complex sample designs which have varying probabilities of selection, would yield biased estimates of the target population and incorrect standard errors (Heeringa et al., 2017; Lumley, 2004)

Let wi denote the survey probability weight for the individual i. Weighted proportions were estimated as:

$$\widehat{\mathrm{P}}_{\mathrm{w}} = \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i} \tag{3}$$

where yi is a binary indicator such as labour force participation or work participation. Weighted means for continuous variables such as wage or earnings were estimated as:

$$\bar{\mathrm{X}}_{\mathrm{w}} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \tag{4}$$

where x_i is the observed income value for an individual i. Weighted standard deviations were used to describe dispersion around the weighted mean.

Descriptive analysis was first used to summarize the distributions of labour force participation, worker-to-population ratio, education level, training status, and employment-type-specific income variables. Categorical variables were reported as weighted percentages, while sample counts were reported as unweighted frequencies. This distinction was maintained because unweighted counts describe the number of observations in the analytical sample, whereas weighted percentages describe the estimated population distribution.

Associations between categorical predictors and binary labour market outcomes were examined using chi-square tests. For a contingency table with observed frequency O_{ij} and expected frequency E_{ij}, the Pearson chi-square statistic is:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij} - E_{ij})^2}{E_{ij}} \tag{5}$$

where r is the number of rows and c is the number of columns. In the survey-weighted analysis, these tests were interpreted using design-adjusted inference. Cramer's V was used to measure the strength of association after chi-square testing. It was computed as:

$$V = \sqrt{\frac{\chi^2}{n(k-1)}} \tag{6}$$

where n is the sample size and k is the smaller value of r or c. Cramer's V was used because statistical significance alone does not describe the magnitude of association. In large samples, even minor differences can be statistically significant. Therefore, a means of interpreting the strength of association was necessary.

The primary multivariable analysis applied binary logistic regression because labour force participation, the dependent variable, was binary. In this instance, the logistic model assessed labour force participation by estimating the log-odds ratio given the levels of education and training, and other variables the model controlled for. The general logistic model is presented as follows:

$$\log\left(\frac{\Pr(LFP_i = 1)}{1 - \Pr(LFP_i = 1)}\right) = \beta_0 + \beta_1 \text{Education}_i + \beta_2 \text{Technical}_i + \beta_3 \text{Training}_i + \beta_k X_i + \varepsilon_i \tag{7}$$

where Pr(LFPi=1) is the probability that an individual i participated in the labour force, Educationi represents education categories, Technicali represents technical education, Trainingi represents vocational or technical training, and Xi represents demographic, social, and location controls. The estimated coefficients were exponentiated and reported as odds ratios:

$$OR = e^{\beta} \tag{8}$$

An odds ratio greater than 1 indicates higher odds of labour force participation relative to the reference group, whereas an odds ratio below 1 indicates lower odds. Statistical significance was evaluated using design-adjusted standard errors and 95% confidence intervals. For an estimated coefficient β, the 95% confidence interval for the odds ratio was calculated as:

$$CI_{95\%} = e^{\hat{\beta} \pm 1.96(SE_{\hat{\beta}})} \tag{9}$$

Four nested logistic regression models were estimated. Model 1 included education level, technical education, and vocational or technical training. This model estimated the unadjusted associations between education and training and labour force participation. Model 2 added demographic controls, including sex, age, age squared, and marital status. Age squared was included to allow a non-linear association between age and labour force participation. Model 3 added social and location controls, including social group, religion and rural–urban sector. Model 4 added State or Union Territory fixed effects to adjust for unobserved geographic differences in labour market structure, education systems, training access, and regional economic conditions.

The model sequence can be written as:

$$\mathrm{M1: LFP_i = f(Education_i, Technical_i, Training_i)} \tag{10}$$

$$\mathrm{M2: M1 + Sex_i + Age_i + Age_i^2 + MaritalStatus_i} \tag{11}$$

$$\mathrm{M3: M2 + Social\ Group_i + Religion_i + Sector_i} \tag{12}$$

$$\mathrm{M4: M3 + State/UT_i} \tag{13}$$

The fourth model was treated as the fully adjusted specification. State or Union Territory fixed effects were included through a set of categorical indicators, with one State or Union Territory omitted as the reference category to avoid perfect multicollinearity. The fixed-effect structure can be represented as:

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta X_i + \sum_{s=1}^{S-1} \delta_s \, \mathrm{State}_{is} + \varepsilon_i \tag{4}$$

where State_{is} is a State or Union Territory indicator and δs is the corresponding fixed-effect parameter.

Regression analysis was performed solely for labour force participation. The worker population ratio and income variables were excluded from the regression modelling process of this study. This was largely to maintain the inferential model's focus on the primary outcome of participation and to avoid intermixing conceptually different dependent variables within one regression. Consequently, income variables were kept for a descriptive comparison across education levels and employment types.

The treatment of incomplete data was through an available-case analysis for each of the statistical methods employed. For the income data, only the valid positive income data were retained, as both a log transformation and meaningful comparisons in income data were only possible with positive income data. During the examination of data for the income analysis, zero income, negative income, and incomplete data were treated as separate cases to avoid an inappropriate log transformation.

The presented sequence of the population-representative, descriptive data analysis led to the assessment of association before the application of the adjusted logistic model. This sequence was also employed in this analysis to determine the association of education and training with participation in the labour force both prior to and following the adjustment for demographic, social, and geographic factors.

3. RESULTS AND DISCUSSION

Table 1 presents the bivariate distribution of the labour force participation rate (LFPR) and the worker-to-population ratio (WPR) across demographic, social, educational, and training characteristics. The estimates show statistically significant variation in LFPR across all reported characteristics, with p-values below 0.001. Because the sample size was large, statistical significance alone was not sufficient for interpretation; therefore, Cramer's V was used to assess the strength of association.

Table 1. Participation Outcomes by Demographic Education and Training Characteristics

Outcome/Characteristic	Category	Unweighted n	Weighted %	χ^2	Cramer's V

LFPR by Sex	Male †	136,789	83.52	41,722.03***	0.391
	Female	135,825	45.20		
LFPR by Sector	Rural †	154,235	67.59	4,087.83 ***	0.122
	Urban	118,379	57.02		
LFPR by Religion	Hindu †	201,312	65.27	1,167.29 ***	0.065
	Muslim	40,784	57.84		
	Christian	19,202	69.54		
	Other religion	11,316	62.80		
LFPR by Social Group	Scheduled Tribe †	39,842	77.63	1,800.67 ***	0.081
	Scheduled Caste	47,909	65.05		
	Other Backward Classes	112,511	63.71		
	Others	72,352	59.83		
LFPR by Education Level	Not literate †	35,430	73.80	10,603.28***	0.197
	Literate / below primary	10,802	77.12		
	Primary	29,730	73.61		
	Middle	63,768	63.23		
	Secondary	46,138	52.01		

	Higher secondary	41,631	48.52		
	Diploma/certificate	4,081	84.33		
	Graduate and above	41,034	70.26		
LFPR by Vocational/ Technical Training	Formal Vocational/ Technical Training	13,026	73.92	48,390.00***	0.421
	Hereditary	25,542	94.50		
	Self-learning	22,720	90.28		
	On-the-job training	26,996	97.75		
	Others	8,383	85.01		
	None †	175,947	49.97		
WPR by Education Level	Not literate †	35,430	73.63	9,649.20 ***	0.188
	Literate / below primary	10,802	76.79		
	Primary	29,730	73.09		
	Middle	63,768	62.17		

	Secondary	46,138	50.96		
	Higher secondary	41,631	46.33		
	Diploma / certificate	4,081	76.85		
	Graduate and above	41,034	60.94		

*Notes: LFPR = labour-force participation rate; WPR = worker-population ratio. Percentages are weighted; sample sizes are unweighted. χ^2, p-value and Cramer's V refer to the overall association between each characteristic and the respective outcome and are therefore reported once per characteristic. * p < 0.05, ** p < 0.01, *** p < 0.001. † denotes the regression reference category. For vocational/technical training, missing or not-asked observations are excluded from the denominator.*

The largest demographic difference was observed by sex. Male LFPR was 83.52%, whereas female LFPR was 45.20%. The chi-square statistic was 41,722.03, with Cramer's V = 0.391, indicating a relatively strong

association between sex and labour force participation. Statistically, this suggests that sex was one of the most important categorical correlates of participation. Socio-economically, the result reflects the continued

gendered structure of work in India, where women's labour market participation is shaped not only by education but also by household responsibilities, marriage patterns, care work, social norms, safety concerns and the availability of acceptable employment. This finding is consistent with earlier research showing that rising education among women in India has not automatically produced proportionate increases in paid employment, partly because education interacts with marriage, household income and social expectations (Chatterjee et al., 2018).

Sectoral differences were also statistically significant. Rural LFPR was 67.59%, compared with 57.02%

in urban areas. The association was weaker than the sex difference but still meaningful, with Cramer's V = 0.122. This rural advantage should not be interpreted simply as better labour market opportunities. In many rural settings, higher participation may reflect economic necessity, family-based work, agriculture, informal self-employment, and limited withdrawal from work. Urban lower LFPR may reflect longer education duration, higher reservation wages, unemployment among educated youth, and social restrictions on women's work.

Religion showed a statistically significant but small association with LFPR (Cramer's V = 0.065). Christian respondents had the highest LFPR at 69.54%, followed by Hindus at 65.27%, other religions at 62.80%, and Muslims at 57.84%. Because the association strength was small, this result should be interpreted carefully. Religious differences in labour market participation may overlap with differences in region, education, gender composition, household structure, and sectoral location. Therefore, the regression analysis was necessary to examine whether these differences remained after adjusting for other predictors.

Social group differences were also statistically significant. ST respondents had the highest LFPR at 77.63%, while Scheduled Caste (SC) respondents had 65.05%, Other Backward Classes had 63.71%, and the "Others" category had

59.83%. The association was small to moderate, with Cramer's V = 0.081. The higher labor force participation rate (LFPR) of Scheduled Tribes (STs) should not be seen as a socio-economic benefit. It could show that STs are more dependent on work in the informal, agricultural, and forest sectors, and withdrawing labor in these sectors is more difficult. Relatively more advantaged groups, on the other hand, have lower participation, and this could reflect longer periods of schooling, effects of household income, or a more deferred entry into work.

The educational pattern was influenced by socio-economic factors and was statistically significant. LFPR was 73.80% among non-literate respondents, 77.12% among respondents who were literate or had education below the primary level, and 73.61% among respondents with primary education. It then decreased to 63.23% for individuals with middle education, 52.01% for secondary education, and 48.52% for higher secondary education. LFPR then increased dramatically to 84.33% among individuals with diploma/certificates, and to 70.26% among graduates and higher. Chi-square statistic was 10,603.28, Cramer's V = 0.197, and it showed a moderate association between education and LFPR.

The patterns indicate a non-linear relationship between education and participation in the labour force. Individuals from lower-education groups had higher participation rates due to early economic participation. Individuals from the Secondary and Higher Secondary groups had lower participation rates, potentially due to the availability of family support, as well as unemployment during the transition period. Higher participation rates were shown by individuals who had attained diplomas and certificates, potentially because qualifications that are diploma- or certificate-based are employment-driven and therefore have a closer relationship with participation in the labour force

This supports the broader human capital argument that education can improve labour market outcomes, but it also agrees with Indian evidence that returns to education vary by level, employment segment and social background (Mitra, 2019; Psacharopoulos & Patrinos, 2018).

Vocational or technical training had the strongest association with LFPR. Respondents with no training had an LFPR of only 49.97%, while LFPR was 73.92% for formal vocational or technical training, 94.50% for hereditary training, 90.28% for self-learning, 97.75% for on-the-job training, and 85.01% for other training. The chi-square statistic was 48,390.00, and Cramer's V = 0.421, the largest association strength reported in Table 1. Statistically, this indicates that training status was more strongly associated with LFPR than education level, sector, religion, or social group. Socio-economically, the result suggests that skill acquisition, especially through work-based training, is closely linked to labour market attachment. However, this should not be interpreted causally, because some training categories, particularly on-the-job training and hereditary training, may be observed among individuals already connected to work. Previous evidence from India also indicates that vocational training is linked to labour market participation and wages, although access and returns vary across gender and sectors (Kumar et al., 2019).

For WPR by education, the pattern was similar to that of LFPR but slightly lower across all education categories. WPR was 73.63% among non-literate respondents, 76.79% among literate or below-primary respondents, 73.09% among primary-educated respondents, 62.17% among middle-educated respondents, 50.96% among secondary-educated respondents, 46.33% among higher-secondary-educated respondents, 76.85% among diploma or certificate holders, and 60.94% among graduates and above. The association between education and WPR was significant, with χ^2 = 9,649.20 and Cramer's V = 0.188. The difference between LFPR and WPR was small for low-education groups but larger for graduates and diploma holders. For example, the graduate LFPR was 70.26%, and the graduate WPR was 60.94%, indicating a difference of nearly 9.32 percentage points, which indicates that more educated groups were more likely to participate in the labour force, but had lower rates of employment during the data collection period. This would be due to participating in economically active job search activities and being selective in the type of employment that they were willing to accept during the period.

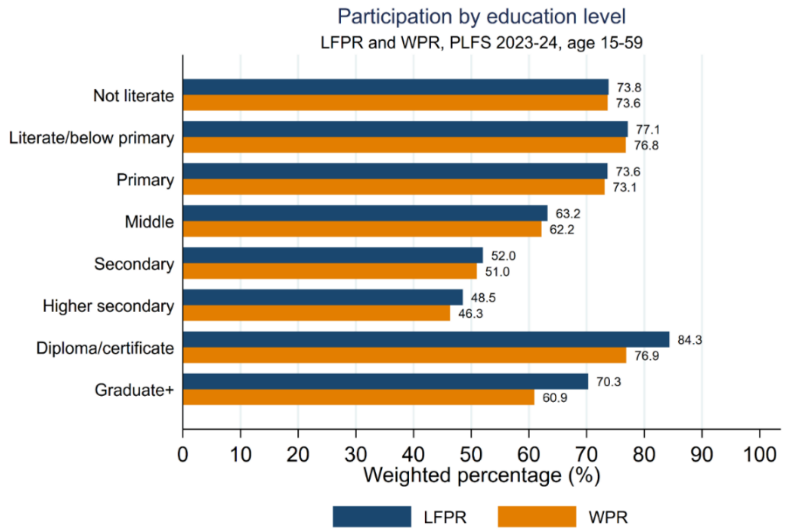

Figure 1. Labour Force Participation and Work Participation by Education Level

Figure 1 illustrates the education non-linearity pattern. It indicates that labour force participation was higher amongst lower-education groups, was lower amongst the Secondary and Higher Secondary groups, and then sharply increased in the Diploma or Certificate groups. The highest LFPR was attained by the Diploma or Certificate group at 84.3%, while the lowest was attained by the Higher Secondary group at 48.5%. The WPR for the two groups also followed the same pattern. Diploma or Certificate holders had a WPR of 76.9%, while Higher Secondary respondents had a WPR of 46.3%.

The figure shows that general education by itself is not enough to produce an upward participation gradient. Instead, labour market enrolment was stronger when education was combined with employable skills, or where individuals had limited capacity for withdrawal from the labour market. In terms of inclusive education, this means that education policy should go beyond the number of school years and incorporate the development of pathways for transitioning from schooling to training, readiness for work, and work placement.

Table 2. Mean Income and Earnings by Education Level and Worker Type

Education Level	Regular Employee		Casual Labour		Self - Employed	
	Percent	Daily Wage/ Salary	Percent	Daily wage	Percent	Daily Gross Earnings
Not literate	5.17	357.44 ±	22.55	367.35 ± 130.	15.71	318.82 ± 262.7

		277.10		85		6
Literate / below primary	2.51	390.82 ± 255.74	8.47	414.54 ± 151.20	5.37	415.39 ± 327.20
Primary	7.64	438.57 ± 306.89	18.41	415.76 ± 160.84	14.43	419.87 ± 310.39
Middle	17.79	499.31 ± 328.25	27.55	439.28 ± 179.58	26.16	465.49 ± 398.19
Secondary	13.75	584.65 ± 426.70	13.50	462.22 ± 193.17	14.82	552.12 ± 462.44
Higher secondary	14.51	694.70 ± 553.77	7.00	444.29 ± 173.99	11.71	578.00 ± 484.63
Diploma / certificate	4.02	952.17 ± 695.44	0.68	615.81 ± 247.26	1.21	814.74 ± 818.83
Graduate and above	34.62	1,389.62 ± 1,093.4	1.83	454.59 ± 177.01	10.58	864.53 ± 906.38

Notes: Values are reported as mean ± standard deviation. Sample sizes are unweighted. Regular employee income, casual labour income, and self-employment income are assessed independently, as each represents a different type of employment income. Casual labour income is interpreted as a daily wage, while self-employment income is interpreted as gross self-employment income. The table only displays positive income values.

Table 2 displays mean income by education attainment level for regular employees, casual employees, and self-employed individuals. It is also statistically necessary to assess these groups separately, as they represent different employment positions. Regular employee income is interpreted as paid employment or salaries, while casual employee income is interpreted as daily paid work, and self-employed income is interpreted as gross self-employment, not paid work.

The education-income gradient was strongest among regular employees. The average regular wage for non-literate employees was 357.44 ± 277.10, while it was 1,389.62 ± 1,093.40 for employees with graduate degrees and higher, meaning that graduate employees earned about four times what non-literate employees earned. Diploma and certificate holders also had relatively high average regular earnings of 952.17 ± 695.44. These findings are in line with human capital theory and with previous studies that showed that education impacts earnings, particularly in wage or salaried employment (Psacharopoulos & Patrinos, 2018; Rani, 2014).

The education gradient was weakest in casual daily employment. For non-literate employees, the average casual daily wage was 367.35 ± 130.85, while it was 454.59 ± 177.01 for employees with graduate degrees and higher. Diploma and certificate holders had the highest casual daily wage, which was also quite low at 615.81 ± 247.26, and considering the small number of casual workers in the group, it is quite insignificant. Relative to regular wage employment, the education gap among casual employees was quite small. Socio-economically, this shows that casual labour markets tend to reward education poorly. This is because casual labour wage rates are set by the demand for physical labour, local labour market conditions, and the casual employee's bargaining and negotiating power. This distinction is necessary because combining all types of earnings into one wage variable would hide the separate impacts that education has in different segments of the labour market. Trivedi and Mishra (2025) also showed regular and casual gender wage differentials and the merits of a statistical decomposition of these wage differentials.

The self-employment earnings trend was also positive, though changes were more subtle across different education levels as compared to employment trends. Mean gross self-employment earnings moved from 318.82 ± 262.76 for individuals with no literacy to 864.53 ± 906.38 for individuals with Graduate and higher qualifications. Individuals with diplomas or certificates had also earned relatively high self-employment income, averaging 814.74 ± 818.83. The high standard deviations, particularly for self-employment income of diploma holders and graduates, suggest that there is a high level of self-employment income inequality. This is expected because self-employment covers a broad range of activities, from low-earning self-employment to high-earning self-employment and self-employment in the form of professional or business activities. From the total number of self-employed people, 60,688 (61.60%) had earned income, whereas 23,754 (27.73%) had no earned income. Income was unknown for 8,810 individuals (10.64%), and only 20 individuals (0.03%) had negative income. Therefore, only positive income was used in the descriptive and log-income analysis since log-transformation of income is not appropriate for zero or negative values. The high number of zero-income earners also shows that the self-employment activities are often seasonal, irregular, and carried out at a household level.

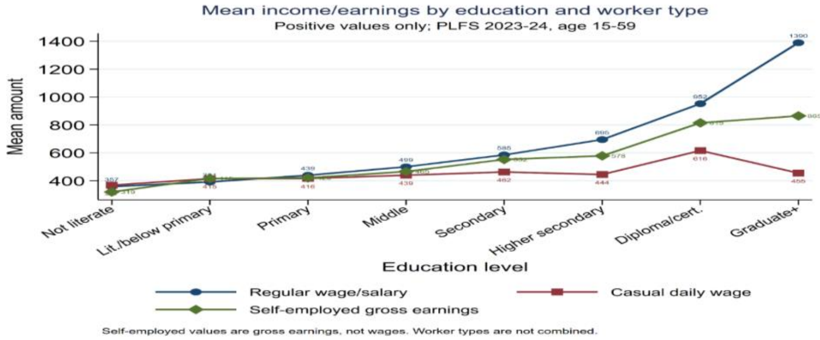

Figure 2. Mean Income and Earnings by Education Level and Worker Type

Figure 2 portrays the same information differently. The regular wage or salary line takes a steep upward trend across education levels, with a particularly large increase from the higher secondary education level to the diploma and graduate education levels. The self-employment line takes a more modest and less steep upward trend with higher education levels. The casual daily wage line is relatively flat across the education levels, showing no substantial graduate premium.

The figure provides a clear socio-economic interpretation: education is most strongly rewarded in regular salaried employment, moderately rewarded in self-employment, and weakly rewarded in casual labour. It can be inferred that education is in varying levels, useful for different jobs, and casual jobs being the least productive. It is necessary to recognise the type of job being carried out along with the level of education. Therefore, for policies that promote inclusive education and transition-to-work, it is important to recognise that the absence of sufficient job opportunities in the formal salaried sector may prevent potential economic gains from the educational advancements.

Four models of logistic regression are presented to analyse the Labour Force Participation Rate (LFPR) in Table 3. Being a binary outcome, Labour Force Participation was coded as 1, and non-participation was coded as 0. The sample included 272,614 persons aged 15-59 years. The estimates are odds ratios, along with their 95% confidence intervals. The analysis was focused solely on LFPR, as the variables of Work Participation Rate and income were excluded from the analysis.

Models differed in their inclusion of variables for education, technical education, and vocational or technical training. Model 2 added demographic variables (sex, age, age squared, and marital status). Model 3 included social group, religion, and work sector. Model 4 included fixed effects of State or Union Territory and was therefore the completely adjusted model. The rest of the models were in a nested form of Model 4. In a statistical sense, the structure of the models was helpful in that, as more relevant demographic, social, and geographic variables were included, the changing relation of education, training, and LFPR was demonstrated.

In the fully adjusted Model 4, technical education remained strongly associated with LFPR. The odds ratio for technical education was 2.011 with a 95% confidence interval of [1.67, 2.42], meaning that individuals with technical education had about twice the odds of labour force participation compared with those without technical education, after adjusting for all included controls. Vocational or technical training showed an even stronger adjusted association, with an odds ratio of 9.094 and a 95% confidence interval of [8.47, 9.77]. This was the strongest regression result in the model. Statistically, the narrow confidence interval and high significance level indicate a precise and robust association. Socio-economically, this reinforces the descriptive finding that training is closely linked with labour market attachment. However, because the analysis is observational, the result should be interpreted as an association rather than causation

Education level showed a more complex adjusted pattern. In Model 4, compared with non-literate respondents, the odds ratio for literate or below-primary respondents was 0.951 and not statistically significant. The odds ratio was 0.893 for primary education, 0.683 for middle education, 0.465 for secondary education, 0.346 for higher secondary education, 0.552 for diploma or certificate education, and 0.710 for graduate and above education. Most education categories had odds ratios below 1 relative to the not-literate reference group. This does not mean education is economically unimportant. Rather, it indicates that after controlling for training, sex, age, marital status, social group, religion, sector, and State or UT, higher education categories were associated with lower odds of being in the labour force than not-literate individuals. This is consistent with the descriptive evidence that low-education groups often participate out of necessity. In contrast, more educated groups may delay labour market entry, continue their education, search longer for suitable jobs, or have higher reservation wages.

Table 3. Survey Design Logistic Regression for Labour Force Participation

Variable	**M1: Education / training**	**M2:+ demographics**	**M3: + social controls**	**M4:+ State/UT FE**
Education Level (ref: not literate)				
Literate / below primary	1.155** [1.05, 1.27]	0.972 [0.87, 1.09]	1.038 [0.92, 1.17]	0.951 [0.85, 1.07]
Primary	0.886*** [0.83, 0.95]	0.801*** [0.74, 0.87]	0.924† [0.85, 1.00]	0.893** [0.82, 0.97]
Middle	0.541*** [0.51, 0.57]	0.603*** [0.56, 0.65]	0.721*** [0.67, 0.78]	0.683*** [0.63, 0.74]
Secondary	0.369*** [0.35, 0.39]	0.397*** [0.37, 0.43]	0.498*** [0.46, 0.54]	0.465*** [0.43, 0.51]
Higher secondary	0.306*** [0.29, 0.33]	0.286*** [0.26, 0.31]	0.368*** [0.34, 0.40]	0.346*** [0.32, 0.38]
Diploma / certificate	0.778* [0.64, 0.95]	0.467*** [0.37, 0.58]	0.605*** [0.48, 0.77]	0.552*** [0.44, 0.70]
Graduate and above	0.665*** [0.62, 0.71]	0.500*** [0.46, 0.55]	0.748*** [0.68, 0.82]	0.710*** [0.65, 0.78]
Training and technical education (ref: no technical education)				
Have Technical education	1.916*** [1.66, 2.21]	1.906*** [1.60, 2.27]	2.126*** [1.76, 2.57]	2.011*** [1.67, 2.42]
Vocational training (ref: no training)				
Have Vocational training	10.600*** [10.02, 11.21]	8.715*** [8.14, 9.33]	8.579*** [8.01, 9.19]	9.094*** [8.47, 9.77]
Demographic controls				

Sex (ref: male)				
Female	—	0.069*** [0.07, 0.07]	0.065*** [0.06, 0.07]	0.063*** [0.06, 0.07]
Age	—	1.523*** [1.50, 1.54]	1.566*** [1.54, 1.59]	1.571*** [1.55, 1.59]
Age squared	—	0.995*** [1.00, 1.00]	0.995*** [0.99, 1.00]	0.995*** [0.99, 1.00]
Marital Status (ref: never married)				
Currently married	—	1.140** [1.05, 1.24]	1.047 [0.96, 1.14]	1.052 [0.96, 1.15]
Widowed	—	1.629*** [1.43, 1.86]	1.551*** [1.36, 1.77]	1.503*** [1.31, 1.72]
Divorced/separated	—	1.683*** [1.33, 2.13]	1.718*** [1.34, 2.21]	1.622*** [1.26, 2.09]
Social Group (ref: Scheduled Tribe)				
Scheduled Caste	—	—	0.424*** [0.38, 0.47]	0.536*** [0.48, 0.60]
Other Backward Classes	—	—	0.429*** [0.39, 0.47]	0.513*** [0.46, 0.57]
Other social group	—	—	0.340*** [0.31, 0.38]	0.409*** [0.37, 0.46]
Religion (ref: Hindu)				
Muslim	—	—	0.714*** [0.67, 0.77]	0.728*** [0.68, 0.79]
Christian	—	—	0.983 [0.87, 1.12]	0.968 [0.83, 1.12]
Other religion	—	—	0.848** [0.76, 0.95]	0.870* [0.76, 1.00]

Sector (ref: rural)				
Urban	—	—	0.491*** [0.47, 0.52]	0.454*** [0.43, 0.48]
State/UT fixed effects	No	No	No	Yes

*Notes: Estimates are odds ratios and are reported with 95% confidence intervals in square brackets. Significance levels: † p < 0.10, * p < 0.05, ** p < 0.01, *** p < 0.001. M4 includes State/UT fixed effects with Jammu & Kashmir as the reference category.*

The demographic controls were also important. In Model 4, the odds ratio for female respondents was 0.063 with a 95% confidence interval of [0.06, 0.07]. This means that women had substantially lower odds of labour force participation than men after accounting for all controls. This result is socio-economically important because it shows that gender differences cannot be explained only by education, training, or location. It indicates that there are barriers on the structural and social levels that restrict women's employment. Indian literature has observed that there are several factors that impact the participation of women in the workforce, namely: education, marriage, care responsibilities, household income, social and cultural beliefs, as well as the availability of jobs (Chatterjee et al., 2018).

Age had a positive association with LFPR, whereas age squared had an odds ratio of just under 1. For Model 4, age had an odds ratio of 1.571, while age squared had an odds ratio of 0.995. This demonstrates a non-linear age-participation relation where workforce participation had a positive association with age, though the association was negative. This type of relationship is expected, as there is typically an increase in workforce participation after school age, participation peaks in the middle adult age, and starts to decline in the later adult age.

Results for marital status were inconsistent. After full adjustment, currently married respondents did not statistically differ from never married respondents (odds ratio = 1.052). However, widows had higher odds of LFPR (odds ratio = 1.503), and divorced or separated respondents also had higher odds of LFPR (odds ratio = 1.622). This may indicate that there is a greater economic need by those who are divorced or separated, or who are widowed, due to a lack of economic support from a spouse. However, this finding should be interpreted cautiously because there is a complex interplay of gender, age, and household structure, which are all brought about by la ack of support from a spouse.

Differences across social groups were still significant in the final model. Social group differences remained statistically significant in the fully adjusted model. Compared with ST respondents, the odds ratios were 0.536 for SC respondents, 0.513 for Other Backward Classes and 0.409 for the "Other social group" category. These results are consistent with the descriptive finding that ST respondents had higher LFPR. However, higher participation among STs may reflect greater livelihood necessity rather than superior labour market outcomes. This is why participation indicators should be read together with employment type and earnings.

In Model 4, Religion still had a significant association. Of the respondents, STs had a lower LFPR, shown in an odds ratio of 0.728. Of the respondents, Christians had no significant difference from Hindus, shown in an odds ratio of 0.968. Other religions had an odds ratio of 0.870, significant at the 5% level. These differences in odds could be due to the unions of other variables like social, regional, gender, and economic. Since the descriptive analysis showed a weak association, the religion variables are to be treated as controlled variables and not primary factors.

Living in urban areas showed a negative association with LFPR. In Model 4, urban areas showed an odds ratio of 0.454, showing a lower LFPR in urban areas in comparison to rural areas, after all other covariates and fixed effects of States or UTs were accounted for. This is consistent with the preliminary analysis in Table 1. This could be a reflection of the urban areas having higher education, higher household earnings, or activities with higher earnings, as well as gender restrictions in the urban market.

The findings demonstrate three main points. First, vocational or technical education was most strongly correlated with labour force participation. Second, more education led to greater earnings among workers and was less closely linked to labour force participation. Third, gender, sector, and social background remained significant even after education and training were accounted for. The findings illustrate that in India, completing the education-to-work transition process is not simply a factor of the level of formal education achieved, but also of pragmatic training, the nature of work, gender, and social positioning.

4. CONCLUSION

This study investigated the correlations between education, vocational and technical training, labour force participation, and earnings (work participation and employment-type-specific earnings) of working-age women and men in India, using the PLFS 2023–24 unit-level data. This research was developed to be a statistics-based analysis of the data, as opposed to a description of social processes lacking empirical foundations. The main innovation of this research was to integrate survey-based, weighted, descriptive statistics, chi-square association with Cramér's V, income comparisons, and logistic regression for labour force participation with four nested survey designs that dealt with the particularities and idiosyncrasies of the Indian context.

The research results indicate that educational attainment cannot solely account for the variations in labor force participation in India. While education and participation are related, the relationship is complex and not straightforward. Individuals in the lower educational categories and those with post-secondary diplomas or certificates are more likely to participate in the labor force. In fact, post-secondary diploma and certificate holders are the most active participants in the labor force, while participants with secondary and higher secondary education have the least participation. It can be assumed that low-educated individuals participate in the labor force to avoid the economic difficulties of remaining unemployed. On the other hand, individuals with secondary or higher secondary education may still be enrolled in educational institutions, or they may be waiting for an opportunity to obtain a job or are facing challenges in transitioning to employment. Vocational and technical education may enhance individuals' employability and participation in the labor force.

Vocational and technical education has the most significant association with a high labor force participation rate (LFPR). It demonstrated the most significant association in a bivariate analysis and in a fully adjusted logistic regression analysis. This participation is significant in the Indian context, especially because of India's large youth demographic and the pressing need for improvement in the educational systems and the steps to improve the labor force participation. Educational institutions must expand the scope of their curricula. The findings indicate that the educational systems of various nations must include practical, technical, and vocational education so as to offer students employment.

The study reinforces gender as an issue of concern in the labor market. Female labor participation significantly lagged behind male participation, and even after accounting for education, training, age, marital status, social group, religion, labor sector, as well as State or UT fixed effects, female labor participation lagged. Thus, women's labor participation lagged more than men's due to more than education and training. Barriers to women's labor participation are structural and socio-cultural. These include the burden of unpaid work, marriage, concerns for safety and inadequate employment opportunities, and work-related socio-cultural barriers. Therefore, policy interventions must embrace education and skills as well as institutional care, secure transportation, flexible employment, and gender-sensitive labor market policies.

Income results reiterate the disadvantage of combining various types of employment in a single wage variable. Regular wage or salary income had a steep positive correlation with education, and the effect was stronger for graduates and diploma holders. Income from self-employment showed a positive but inconsistent correlation with education, while casual employment showed a weaker positive correlation. This is a significant statistical and conceptual input in the study. Merging all types of income would lead to the loss of the positive correlation of education with formal employment, casual employment, and self-employment. For the Indian context, this is particularly important as a significant portion of the workforce is informal and self-employed. For the balance of labor research and earnings in developing economies, the study is a particularly valuable framework as it maintains the diverse employment models of developing economies.

This research has a tripartite statistical innovation. Firstly, surveys are integrated with PLFS design-consistent analysis and survey probability weights. Thus, the estimates are more relevant and appropriate for population-level estimates. Secondly, measures of association strengths are integrated with some regression models, and the interpretation of these models is not limited to p-values. Thirdly, the study uses four nested logistic regression models of LFPR. The first model includes education and training. In subsequent models, demographic and social variables are added, followed by fixed effects of the States and UTs of India. The sequence of the models illustrates the impact of different control variables on education and training associations. Such a design provides readers with the opportunity to see whether findings are stable across different models, thus improving the overall quality and transparency of the research.

The research study builds the case for evidence-based articulation of education and skills policies, along with elements of labour market planning and transition-to-work policies in India. The first step of policy formulation, increased enrolments and educational attainments, is necessary, but not sufficient, for providing a more optimal educational framework. The study also highlights the need for more support for women, socially disadvantaged groups, and people who seek jobs in urban areas, as well as people who are moving from general education to the job market.

The study has some value internationally as many developing countries have large youth populations, large informal labor markets, educational systems that are expanding rapidly, gender gaps, and inconsistent returns on education. The Indian case shows that in the statistical analysis of education and labor outcomes, participation, employment, and regular wages should be analyzed separately from casual and self-employment wages. Other countries that have labor force survey data can implement this analysis.

The study suggests that education, training, gender, sector, and social background affect labor market m participation in India. Technical and vocational training are very important for the participation of workers in the labor force. Education is more highly valued in the formal labor market and in regular employment than in casual employment. This shows that inclusive educational policy should consider the connection to the labor market and the skills needed for the job. India and other countries can benefit from creating labor and educational policies that are focused on the intersection of education and the labor market.

Acknowledgements

No funding, institutional support, or external assistance was received for this study.

Author Declaration

The authors declare no conflicts of interest and that this manuscript is original and has not been published elsewhere. The manuscript is not under consideration by another journal.

References

Aakanksha & Bishnoi, N. K. (2024). Education and earnings: Examining the returns across regular, casual, and self-employed labour markets. SN Business & Economics, 4, Article 64. https://doi.org/10.1007/s43546-024-00663-0

Agrawal, T. (2012). Vocational education and training in India: Challenges, status and labour market outcomes. Journal of Vocational Education & Training, 65(4), 453–474. https://doi.org/10.1080/13636820.2013.834954

Agrawal, T. (2014). Gender and caste based wage discrimination in India: Some recent evidence. Journal for Labour Market Research, 47(4), 329–340. https://doi.org/10.1007/s12651-013-0152-z

Agrawal, T., & Agrawal, A. (2017). Vocational education and training in India: A labour market perspective. Journal of Vocational Education & Training, 69(2), 246–265. https://doi.org/10.1080/13636820.2017.1303785

Ahmed, T. (2016). Labour market outcome for formal vocational education and training in India: Safety net and beyond. IIMB Management Review, 28(2), 98–110. https://doi.org/10.1016/j.iimb.2016.05.002

Bairagya, I., Bhattacharya, T., & Tiwari, P. (2021). Does vocational training promote female labour force participation? An analysis for India. Margin: The Journal of Applied Economic Research, 15(1), 149–173. https://doi.org/10.1177/0973801020976601

Chatterjee, E., & Vanneman, R. D. (2022). Women's low employment rates in India: Cultural and structural explanations. Population and Development Review, 48(2), 445–474. https://doi.org/10.1111/padr.12474

Chatterjee, E., Desai, S., & Vanneman, R. (2018). Indian paradox: Rising education, declining women's employment. Demographic Research, 38, 855–878. https://doi.org/10.4054/DemRes.2018.38.31

Chen, J., Kanjilal Bhaduri, S., & Pastore, F. (2022). Updates on returns to education in India: Analysis using PLFS 2018 19 data (GLO Discussion Paper No. 1016). Global Labor Organization. https://doi.org/10.2139/ssrn.4114527

Choudhury, P. K., Mallick, H., & Kumar, A. (2024). Vocational education, skill training and self employment: Evidence from India's non farm sector. Journal of Vocational Education & Training. https://doi.org/10.1080/13636820.2024.2359369

Desai, S., & Joshi, O. (2019). The paradox of declining female work participation in an era of economic growth. The Indian Journal of Labour Economics, 62(1), 55–71. https://doi.org/10.1007/s41027-019-00162-z

Heeringa, S. G., West, B. T., & Berglund, P. A. (2017). Applied survey data analysis (2nd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781315153278

Hussain, Y. R., & Mukhopadhyay, P. (2023). How much do education, experience, and social networks impact earnings in India? A panel data analysis disaggregated by class, gender, caste and religion. SAGE Open, 13(4). https://doi.org/10.1177/21582440231220942

Klasen, S., Le, T. T. N., Pieters, J., & Santos Silva, M. (2021). What drives female labour force participation? Comparable micro-level evidence from eight developing and emerging economies. The Journal of Development Studies, 57(3), 417–442. https://doi.org/10.1080/00220388.2020.1790533

Kumar, R., Mandava, S., & Gopanapalli, V. S. (2019). Vocational training in India: Determinants of participation and effect on wages. Empirical Research in Vocational Education and Training, 11, Article 3. https://doi.org/10.1186/s40461-019-0078-y

Lumley, T. (2004). Analysis of complex survey samples. Journal of Statistical Software, 9(8), 1–19. https://doi.org/10.18637/jss.v009.i08

Mishra, S., & Trivedi, S. (2024). Gender wage disparity in urban India: An analysis using unconditional quantile regression. Demography India, 53(2), 25–46. https://iasp.ac.in/uploads/journal/3-1737717927.pdf

Mitra, A. (2019). Returns to education in India: Capturing the heterogeneity. Asia & the Pacific Policy Studies, 6(2), 151–169. https://doi.org/10.1002/app5.271

National Statistical Office. (2024). Periodic Labour Force Survey annual report July 2023–June 2024. Ministry of Statistics and Programme Implementation, Government of India. https://www.mospi.gov.in/sites/default/files/publication_reports/AnnualReport_PLFS2023-24L2.pdf

Psacharopoulos, G., & Patrinos, H. A. (2018). Returns to investment in education: A decennial review of the global literature. Education Economics, 26(5), 445–458. https://doi.org/10.1080/09645292.2018.1484426

Rani, P. G. (2014). Disparities in earnings and education in India. Cogent Economics & Finance, 2(1), Article 941510. https://doi.org/10.1080/23322039.2014.941510

Reed, M. N. (2020). The labor force participation of Indian women before and after widowhood. Demographic Research, 43, 673–706. https://doi.org/10.4054/DemRes.2020.43.24

Refeque, M., & Azad, P. (2022). How do linguistic and technical skills affect earnings in India? Indian Economic Review, 57(1), 23–57. https://doi.org/10.1007/s41775-022-00132-1

Schneider, S. (2024). India's labour market challenges: Employability of young workforce from the perspective of supply and demand. PROSPECTS. https://doi.org/10.1007/s11125-024-09691-y

StataCorp. (2013). Stata survey data reference manual: Release 13. Stata Press. https://www.stata.com/manuals13/svy.pdf

Tripney, J. S., & Hombrados, J. G. (2013). Technical and vocational education and training for young people in low and middle income countries: A systematic review and meta analysis. Empirical Research in Vocational Education and Training, 5, Article 3. https://doi.org/10.1186/1877-6345-5-3

Trivedi, S., & Mishra, S. (2025). A quantile regression modelling approach to study gender wage gap in India. Statistika: Statistics and Economy Journal, 105(1), 102–122. https://doi.org/10.54694/stat.2024.18