Returns to vocational education and training in India:
Evidence from Periodic Labour Force Survey 2023-24

Mentor: Prof. Saudamini Das

Submitted by Vishnu K Venugopal

Abstract

The development of skilled labour resources in a nation depends heavily on vocational education and training, or VET. The Indian government makes large investments in the nation's skilling ecosystem. This study examines the effects of various vocational education and training programs on the daily incomes of self-employed people and regular or salaried employees. For this study, the Periodic Labour Force Survey (PLFS) 2023–2024 serves as the primary data source. Both the ordinary least squares method and the quasi-experimental propensity score matching method are used in this study to determine how various forms of vocational education and training affect the incomes of regular or salaried employees as well as self-employed individuals. As compared to informal vocational education and training or no vocational education and training, the study finds that formal vocational education and training yields a significantly higher earnings premium. The impact of different types of vocational education and training on the earnings of self-employed individuals is either marginal or not statistically significant.

Keywords: Vocational Education and Training, Propensity Score Matching, Employment, Wages, On-the-job training.

JEL Classification: I26, J21, J24, J26, J31, O15

1. Introduction

India is at a crucial turning point where its young and growing workforce presents a unique opportunity to leverage the demographic dividend. However, simply having a large working-age population is not enough; individuals must be equipped with the right skills to turn this potential into economic growth and social advancement. Skill development is vital for improving employability, encouraging entrepreneurship, and boosting industrial productivity, ensuring that India's youth can meet the changing demands of a dynamic job market.

Vocational Education and Training (VET) is a key component of the skill development ecosystem in any country. VET is defined as "the learning that aims to acquire knowledge, know-how, information, values, skills, and competencies – whether job-specific or transversal – required in specific occupations or more broadly in the labour market" (Cedefop, 2021). Vocational education is often conceived as a crucial element in the strategy for enhancing the competitiveness of economies (Van den Berg et.al 2006). Further, it is also considered as a solution to the problem of unemployment especially youth unemployment (Doerr et. al 2017).

In the initial decades following India's independence, there was a strong emphasis on higher education, while vocational education was largely neglected (Tilak, 2003). Vocational training was primarily provided through Industrial Training Institutes (ITIs) and Industrial Training Centres (ITCs). A significant boost to vocational education in India came with the "Vocationalisation of Education" scheme introduced in 1988, along with its revised version in the Ninth Five-Year Plan. This initiative led to the introduction of vocational courses at the XI and XII levels. With the Eleventh Five-Year Plan and the subsequent National Skills Policy and National Skill Development Mission, vocational education and skill development began to receive increased attention from policymakers.

The official estimates based on the NSSO Employment Unemployment Survey and Periodic Labour Force Survey (PLFS) convey that the proportion of formally trained has increased from 2.4% in 2004-05 to 4.1% in 2023-24. This increase in the proportion of formally trained would also be attributable to the enhanced efforts on the part of the government in strengthening the skilling ecosystem. The extent to which the efforts towards skilling the population will succeed in addressing the issue of youth unemployment and in enhancing the earnings potential of workers will depend also on the structure of the labour market and earnings differentials across sectors.

This paper analyzes the returns on different types of vocational education and training in India. This will help in understanding the returns on investments made by the government in vocational education and training.

2. Review of Literature

The studies on vocational education and training have explored its various dimensions such as conceptualization of vocational education and training in a larger theoretical framework, factors determining participation in vocational education, and returns to vocational education in terms of labour market outcomes.

Definition of vocational education

Vocational education is a concept that can be difficult to clearly define. According to Moodie (2002), understanding the identity of vocational education involves considering four key characteristics: epistemological, teleological, hierarchical, and pragmatic. From an epistemological perspective, vocational training is primarily associated with practical skills and applied knowledge, in contrast to theoretical learning. The teleological aspect highlights the role of vocational education in preparing individuals for employment, differentiating it from general education, which focuses more on intellectual development and civic engagement. The hierarchical characteristic links vocational education to specific occupational levels, often placing it between secondary and higher education. Lastly, the pragmatic aspect contributes to its adaptability and is frequently defined by what it is not. Taking into account these multidimensional characteristics, vocational education can be broadly defined as the development and application of knowledge and skills for mid-level occupations that society needs at various times (Moodie, 2002).

The definitions of vocational education and training provided by various agencies can be seen as more or less in line with the definition provided by Moodie (2002). Vocational education and training involve acquiring knowledge, skills, and competencies necessary for specific occupations or more broadly in the labour market (European Commission and Cedefop, 2021) The focus of this study is on vocational education and training in general which is imparted at the post-secondary level and beyond.

Theoretical framework

The research on vocational education and training in economics has mainly framed it within human capital theory. The human capital model suggests that individuals make rational choices about investing in education and training. Their goal is to weigh the direct costs and lost wages against the future benefits that education or training will provide. These models emphasize the importance of having information about expected wages, while also acknowledging that non-monetary factors play a significant role (Becker, 1993). This means that, all else being equal, the demand for education tends to be stronger when individuals anticipate benefits that will last for a longer time and when the discount rate is relatively low. Studies within the human capital framework also try to link investments in human capital (education and training) with broader economic growth. Schultz (1963) expands on human capital theory by linking education, including vocational training, to economic growth. He also emphasizes the role of government in financing education and training, as markets may underinvest in human capital.

One major problem with human capital theory is its insufficient focus on how students develop expectations regarding future labour market conditions. The decisions regarding human capital investment often rely on factors beyond financial rewards, such as insights into labour market conditions and the future potential of various educational pathways. (Borghans et al., 1996). In such a situation, it is desirable policy to publish detailed medium forecasts of labour market positions in various vocations so as to help the students in their decision-making (Borghans et al., 1996).

Participation in vocational training

The factors influencing an individual's decision to participate in vocational education and training are wide-ranging, encompassing individual and household characteristics as well as the associated costs and benefits.

There has been no consensus among studies regarding the influence of general education on the decision to participate in vocational education. Some studies have noted that those whom didn't perform well in general education streams are opting for vocational education. VET is seen as an educational system for the economically disadvantaged and those who are educationally underprivileged, who may not qualify for higher education admission (Tilak, 2003). But Aypay (2003) in the context of Turkey and Moenjak & Worswick (2003) in the context of Thailand have found that academic achievement is positively and significantly related to the choice of vocational education.

In the Indian context, studies have shown that variables such as age, education and marital status do play a role in the choice of formal vocational education and training (Kumar, Mandava and Gopanappalli, 2019; Anikin, 2021). These studies use logistic regression and multinomial logistic regression to understand the factors that influence the choice into formal and informal vocational training.

Returns to vocational educational and training

The comparison between the returns of vocational education and general education has always been a topic of debate.

Many studies have noted that the returns to vocational education are lower than that of general education (Psacharapoulos, 1987; Psacharapoulos, 2024). Psacharapoulos and Patrinos (1993) finds that, in the content of Latin American countries, vocational schooling fares better than general secondary schooling in terms of private returns. But when cost of schooling is taken into consideration, and when social returns are estimated, the returns to vocational education are much lower. The evidence of lower returns to vocational education vis-a-vis general education inspired a reduction in World Bank funding for vocational education. This was also part of World Bank's shift in focus towards poverty reduction and their faith in the potential of primary education to reduce poverty (Bennell and Segerstrom, 1998). However, there is a renewed interest in vocational education across various countries and international institutions such as UNESCO (Debroy, 2009).

Many studies in the context of developing countries have reported positive and significant returns to vocational education. Vocation education in Thailand generated higher returns compared to general education after controlling for the selection bias (Moenjak and Worswick, 2003). In a study conducted in Egypt using the Egyptian Household Survey, El-Hamidi (2006) finds that men with vocational education receive higher returns compared to those with only general secondary education. Tunali (2002) finds that vocational education yields heterogeneous returns across genders in Turkey, benefiting women through a higher likelihood of employment and helping men earn better wages.

The returns to vocational education have primarily been examined through two methods: rate of return analysis and regression analysis. Rate of return analysis focuses on the flow of returns and costs related to vocational education (Psacharopoulos, 1995). This method was augmented to assess the social rate of return by incorporating the subsidy costs incurred by the government and the taxes paid by individuals while employed in the analysis (Tilak, 1987). The studies which use regression analysis often use the Mincerian earnings equation (Mincer, 1974) which is a semi-logarithmic earnings function to estimate the returns to education in general and vocational education in particular (Duraisamy, 2002).

The problem with using a regression analysis to estimate the returns to vocational education is that of the “selection bias” or “ability bias” which is not considered in the Mincerian earnings equation (Ahmed and Chattopadhyay, 2016). The decision to pursue vocational education instead of general education may depend on an individual's talents or abilities. This can create endogeneity problems in the regression analysis. Many studies use the Heckman selection model with a two-step estimation technique to account for self-selection in the labour market (Ahmed and Chattopadhyay, 2016; Banerjee, 2017; Bahl, Bhatt and Sharma, 2021; Vincent and Rajsekhar, 2023). The Heckman selection model requires the inclusion of exclusion variables in the selection function. While using an instrumental variable can help address the problem of endogeneity, it has been noted in studies on the returns to education that finding an appropriate instrumental variable within the dataset is extremely challenging (Carneiro et al., 2003).

Efforts have also been made to utilize the underlying heteroscedasticity within the model to provide valid instruments for addressing endogeneity and estimating the causal impact of vocational education on earnings (Bahl, Bhatt, and Sharma, 2021). Some studies have utilized the quasi-experimental method of propensity score matching, which employs observable characteristics of individuals to create a counterfactual. This approach allows researchers to estimate the causal effect of vocational education on wages (Tognatta, 2014; Bahl, Bhatt, and Sharma, 2021).

3. Research Question and Methodology

3.1 Theoretical Framework

Most theoretical models of investment in education and training have been developed within economic, sociological, or integrated frameworks. Economic models, particularly the human capital theory (Becker, 1962; Schultz, 1961), have been widely used in research on educational decision-making since their emergence in the 1960s. The human capital model suggests that individuals or households make rational decisions about investing in education and training by weighing direct costs and lost earnings against the anticipated benefits of such investments. The present study also adopts a human capital theory framework wherein the private returns to vocational education and training accrue to the individuals in the form of higher earnings. Evaluating investments in education and training presents a key challenge, particularly in assessing the benefits perceived by individuals when making educational decisions. Ideally, individuals' subjective expectations regarding the returns on education and training would provide valuable insights. However, such perceptions are often difficult to measure due to their intangible and individualized nature. Drawing on Billett (1998), this framework assumes that actual economic returns serve as reliable measures of the efficiency of education and training systems. Even in the absence of data on individual perceptions, these real-world outcomes provide a basis for evaluating the effectiveness of educational investments, aligning with human capital theory's emphasis on rational decision-making in education and workforce participation.

3.2 Research Gap

- Previous studies have not included the earnings of the self-employed in analysing the returns to vocational education and training.
- There have been no studies conducted regarding the post-COVID period.

3.3 Objective

- To examine the earnings premium associated with formal and informal vocational education and training compared to no vocational education and training
- To explore the differential impact of vocational education and training by type of employment.

3.4 Research Question

1. What is the impact of different types of vocational education and training on daily earnings?

    1.1 What is the impact of formal vocational education and training compared to informal vocational education and training on daily earnings?

    1.2 What is the impact of formal vocational education and training compared to lack of vocational education and training on daily earnings?

    1.3 What is the impact of informal vocational education and training compared to lack of vocational education and training on daily earnings?

2. What is the differential impact of vocational education and training on the earnings of self-employed and earnings of regular or salaried employees?

3.5 Methodology

This paper attempts to understand the returns to vocational education and training using the methods of the ordinary least squares (OLS) model and the quasi-experimental method of propensity score matching (PSM).

In the Ordinary Least Squares method, a Mincerian earnings equation is estimated to understand the impact of the key treatment variable (type of vocational education) on daily earnings. The equation that will be estimated is as follows:

$\ln E_i = \beta 0 + \beta 1 VET_i + \beta 2 X_i + u_i$

where the dependent variable is the logarithm of daily earnings, VET_i is the dummy for formal VET, informal VET, and no VET, X_i is the set of other control variables that can influence the daily earnings and u_i is the error term. The set of control variables are human capital variables like years of formal education, age, and age squared as well as personal and household characteristics such as religious group, social group, gender, sector(rural/urban), and marital status. The dummy variables for the region[1] are also included to control for unobserved regional characteristics that may influence earnings.

[1] North: Jammu and Kashmir, Himachal Pradesh, Punjab, Uttarakhand, Haryana, Uttar Pradesh, Bihar, Chandigarh, Delhi; East: West Bengal, Jharkhand, Odisha, Chattisgarh; West: Rajasathan, Madhya Pradesh, Gujarat, Maharashtra,

The choice of vocational education and training can be an endogenous variable in the regression because of the selection bias involved. The individuals who choose vocational education might be systematically different from those who don't. Thus, comparing a group with vocational education with those without vocational education might lead to biased estimates. The method of propensity score matching tries to address this problem by creating a statistical counterfactual based on observable characteristics for individuals with vocational education and training while those with no vocational education and training serve as the control group.

The propensity score, denoted as p(X), is estimated using a logistic regression model:

$$p(X) = P(T_i = 1 \mid X_i) = \frac{\exp(\beta_0 + \beta_1 X_i)}{1 + \exp(\beta_0 + \beta_1 X_i)}$$

where:

- T_i is the treatment indicator (1 if individual i received formal vocational education and training, 0 otherwise),
- X_i is the vector of covariates,
- β are the estimated coefficients.

After estimating p(X), treated individuals are matched with untreated individuals based on their propensity scores using the caliper matching technique[2].

Once matching is performed, the Average Treatment Effect on the Treated (ATT) is estimated as follows:

Goa; North East: Sikkim, Arunachal Pradesh, Nagaland, Manipur, Mizoram, Tripura, Meghalaya, Assam; South: Andhra Pradesh, Karnataka, Kerala, Tamil Nadu, Telengana, Puducherry, Andaman and Nicobar Islands, Lakshadweep, Dadra and Nagar Haveli, Daman and Diu.

[2] A treated unit is matched only with a control unit whose propensity score is within a specified range (caliper). Caliper of 0.05 is used in the models that are estimated in this study.

The Average Treatment Effect on the Treated (ATT) is given by:

$$ATT = \mathbb{E}[Y(1) - Y(0) \mid T = 1] = \frac{1}{N_T} \sum_{i \in T} \left(\ln Y_i - \sum_{j \in C(i)} w_{ij} \ln Y_j \right)$$

where:

- Y_i is the daily earnings of individual i,
- $C(i)$ represents the set of matched control units for treated individual i,
- w_{ij} are the weights assigned in the matching process,
- N_T is the number of treated individuals.

There are three key assumptions used in the propensity score matching technique. The first assumption is conditional independence, which implies that an individual's selection into the treatment group is based solely on observable characteristics and not influenced by unobservable factors common to both the treatment and control groups. The second assumption is the common support condition, meaning that for any given set of observable characteristics, there are sufficient observations in the control group that resemble those in the treatment group but did not receive the treatment. In other words, observable characteristics should not fully determine whether an individual opts for vocational education. The third assumption ensures no spillover effects between the treatment and control groups. This means that the outcome variable for an individual should remain unaffected by the presence or concentration of treated or untreated individuals within the sample.

3.6 Data Source

The data source used in the study is Periodic Labour Force Survey 2023-24. The data was collected during July 2023- June 2024. The Survey covers the whole of the Indian Union except the villages in Andaman and Nicobar Islands. The Survey uses a stratified multi-stage design. The first stage units (FSU) were the latest available Urban Frame Survey (UFS) blocks in urban areas and 2011 Population Census villages (Panchayat wards for Kerala) in rural areas. The ultimate stage units were households. In the case of large FSUs one intermediate stage unit, called hamlet group/ sub-block was formed.

Since survey data is used for the study and different households have different probabilities of being selected into the sample, weights have to be used to correct for the sample design

(Deaton,1997). The OLS model is run using survey weights. Further in rural areas, the clusters are often villages and the observations from one cluster are much more like one another than are observations from different clusters (Deaton, 1997). Therefore, robust standard errors which are clustered around the first-stage units (FSU) are used in the OLS models.

4. **Results**

4.1. **Descriptive Statistics**

The nature of vocational training received will have a bearing on an individual's labour market status and employment status. The direction of causality might be in either direction. The individuals who received certain types of training might prefer certain forms of employment and those in certain forms of employment might prefer to have certain types of training. The percentage of individuals in different employment statuses based on the type of training they received is shown in Table 4.1.

Table 4.1: Distribution of workers by economic activity status and VET

Whether received any Vocational/Technical Training	Employment Category Status (Current Weekly Status)						
	Self employed	Regular or salaried Employees	Casual labour	Unemployed	Not in Labour Force	Unemployment Rate	Total
Formal	23.39	37.23	1.62	9.21	28.55	12.89	100
Hereditary	71.32	2.65	11.7	1.03	13.31	1.19	100
Self-learning	56.53	11.85	16.47	1.78	13.36	2.05	100
Learning on the job	27.86	46.23	19.83	2.07	4.02	2.16	100
Others	50.02	18.01	10.84	2.9	18.24	3.55	100
No training	22.5	10.45	9.98	3.48	53.59	7.50	100
Total	31.83	14.27	11.24	3.16	39.51	5.22	100

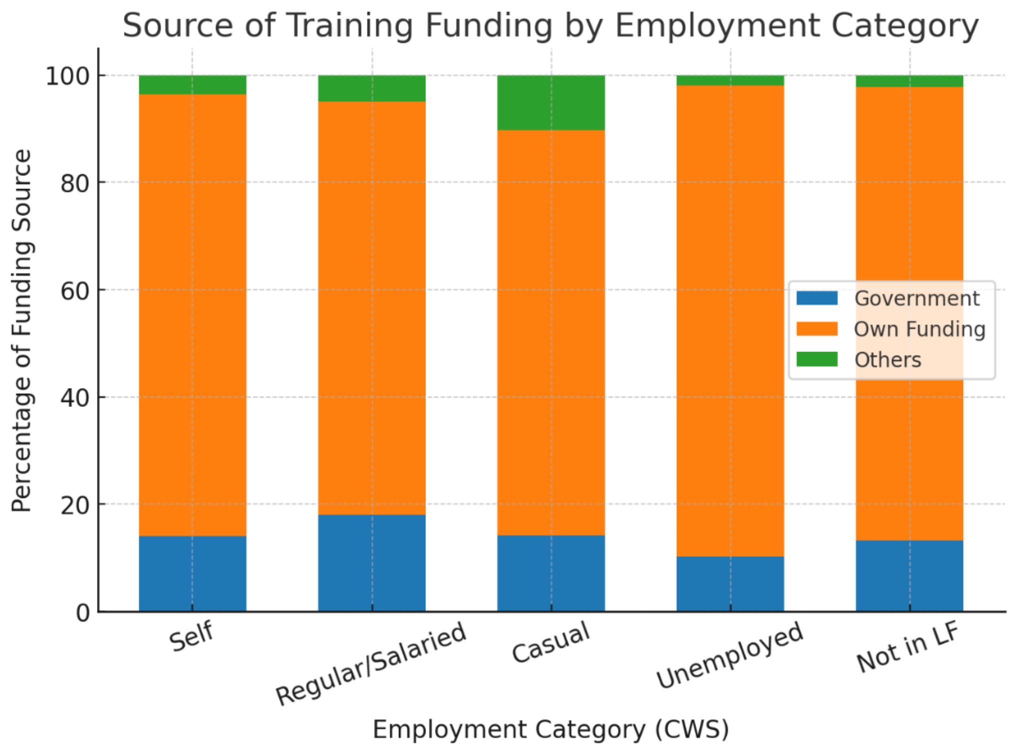
The individuals who are in regular or salaried employment have a significant proportion among those who received formal training and "learning on the job" type of informal training. The proportion of self-employed individuals is higher among those who received various forms of informal training, such as hereditary training, self-learning, and other informal methods compared to those who received formal training and "learning on the job" type of informal training. The prevalence of casual employment also seems to be relatively higher among those who received informal training compared to those who received formal training. One important fact to be noted is that there is a significant proportion of individuals who are not in the labour force (53.5%) among those who haven't received any form of training. The unemployment rate seems to be much higher among formally trained compared to those who received informal forms of training and those who didn't receive any training.

The fact that a significant proportion of individuals remain out of the labour force despite receiving different forms of vocational training seems puzzling. But an exploration into the gender-wise data shows that this paradox is caused by the high share of formally and informally trained women who are out of the labour force (see Table A1 in the appendix). It can be seen that 46.5% of the women who received formal training and out of the labour force.

The relationship between formal training and employment status varies based on the type of formal training. The individuals who have received 'on the job' formal training is employed in high percentages in regular or salaried employment. This hold holds true for males (77.03%) as well as females (80.24%). The proportion of regular or salaried employees is lower among those who received part-time formal training and full-time formal training compared to 'on the job' formal training. This proportion is much lower among females compared to males.

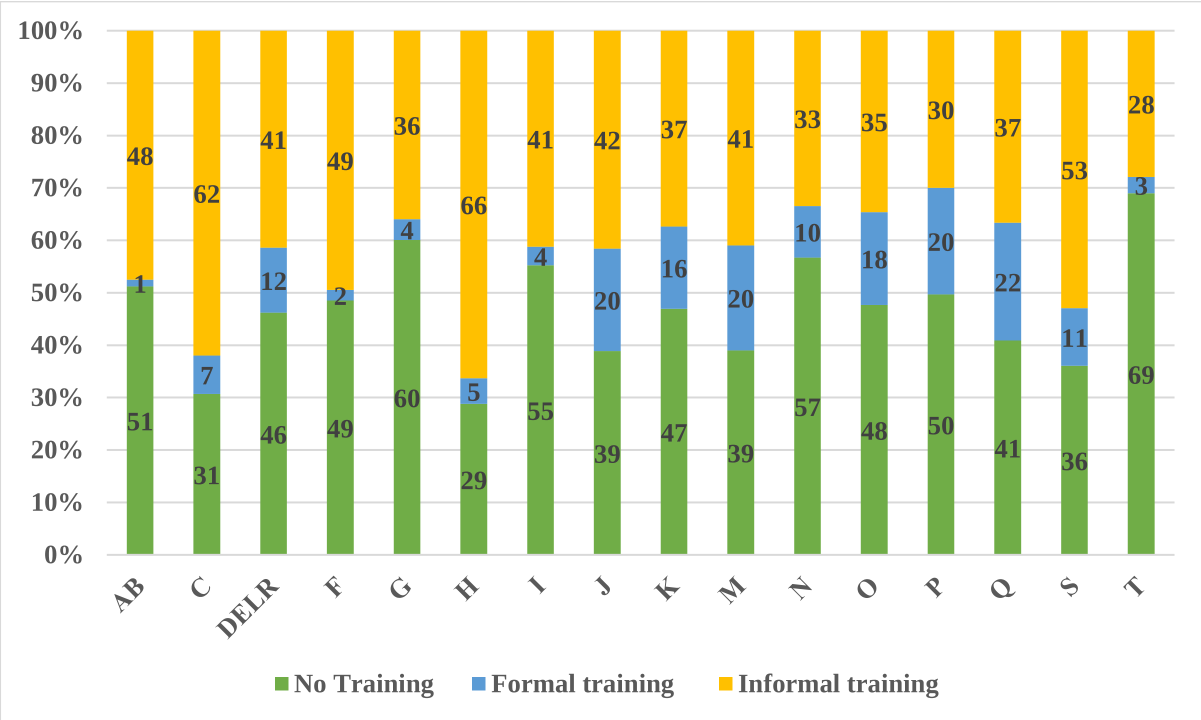
The employment status of an individual may also be related to the duration of formal training that he/she has taken. It can be seen that the proportion of unemployed is lowest (4.8%) among individuals who have received formal training of the shortest duration (less than 3 months). The proportion of unemployed among formally trained seems to vary positively with the duration of training and seems to be highest among those who received training beyond 18 months. Another interesting fact to note is that among those who received formal training, the proportion of those who are not in labour force seems to be lowest among those who have a duration of training beyond 18 months. Thus, it means that those who received training of high duration i.e., beyond 18 months are more likely to be in regular employment or remain unemployed rather than staying out of the labour force. The proportion of self-employed and casual employed individuals does not vary significantly with the duration of formal training.

Figure 4.1: Distribution of workers by economic activity status and source of VET funding



The source of funding for training significantly influences the decision to participate in training. The figure shows that among all categories of employment, own funding is the most preferred option for securing formal training. Among all categories of employment, the share of those who rely on government funding for formal training is much less compared to own funding, but much higher than other sources of funding.

Figure 4.2: Distribution of workers by sector of employment and status of VET

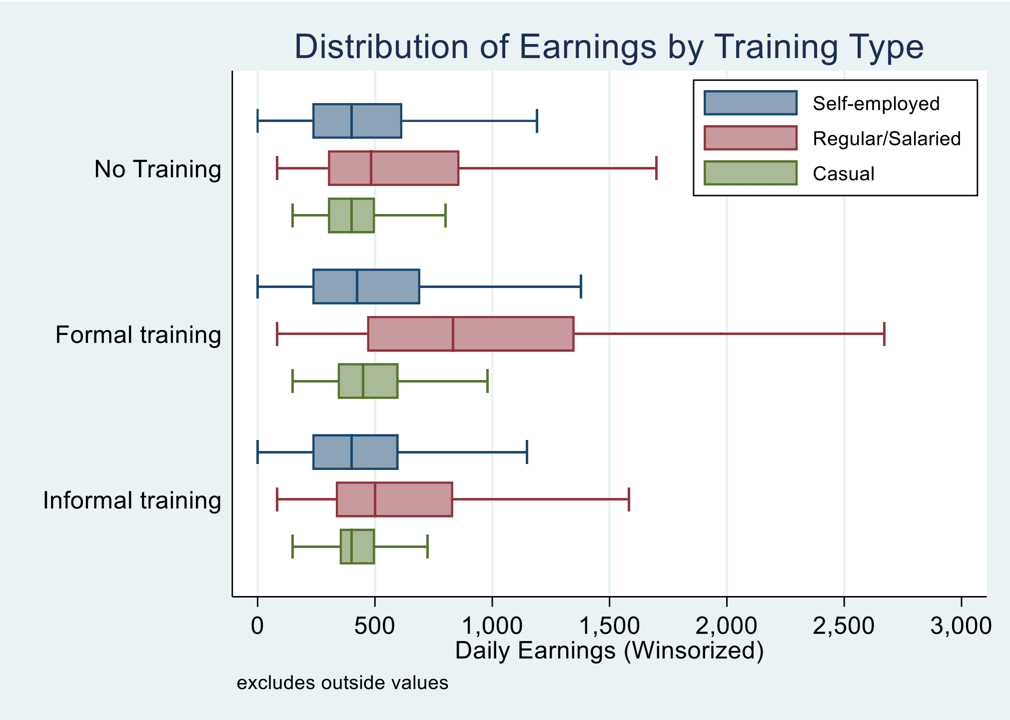


AB: Agriculture, forestry and fishing, mining and quarrying; C: Manufacturing; DELR: Electricity, gas, water supply, sewerage, real estate activities and arts, entertainment and recreation; F: Construction; G: Wholesale and retail; H: Transportation and storage; I: Accommodation and food; J: Information & communication; K: Financial and insurance activities; M: Professional, scientific and technical activities; N: Administrative and support service activities; O: Public administration and defence; P: Education; Q: Human health and social work; S: Other service activities; T: Activities of households as employers

The chart illustrates the distribution of No Training, Formal Training, and Informal Training across various industries, revealing significant disparities. A substantial proportion of the workforce in several sectors has not received any training, with the highest percentage observed in Activities of Households as Employers (69%), followed by Construction (60%) and Agriculture, Forestry & Fishing (51%). These industries, often characterized by informal employment and low entry barriers, may rely on traditional knowledge and experiential learning rather than structured skill development. Informal training, which typically occurs on the job without standardized certification, is highly prevalent in Manufacturing (62%) and Transportation & Storage (66%), where practical skills are often acquired through experience rather than formal education. This suggests that industries with a strong manual labor component depend heavily on informal skill transfer rather than institutionalized training programs.

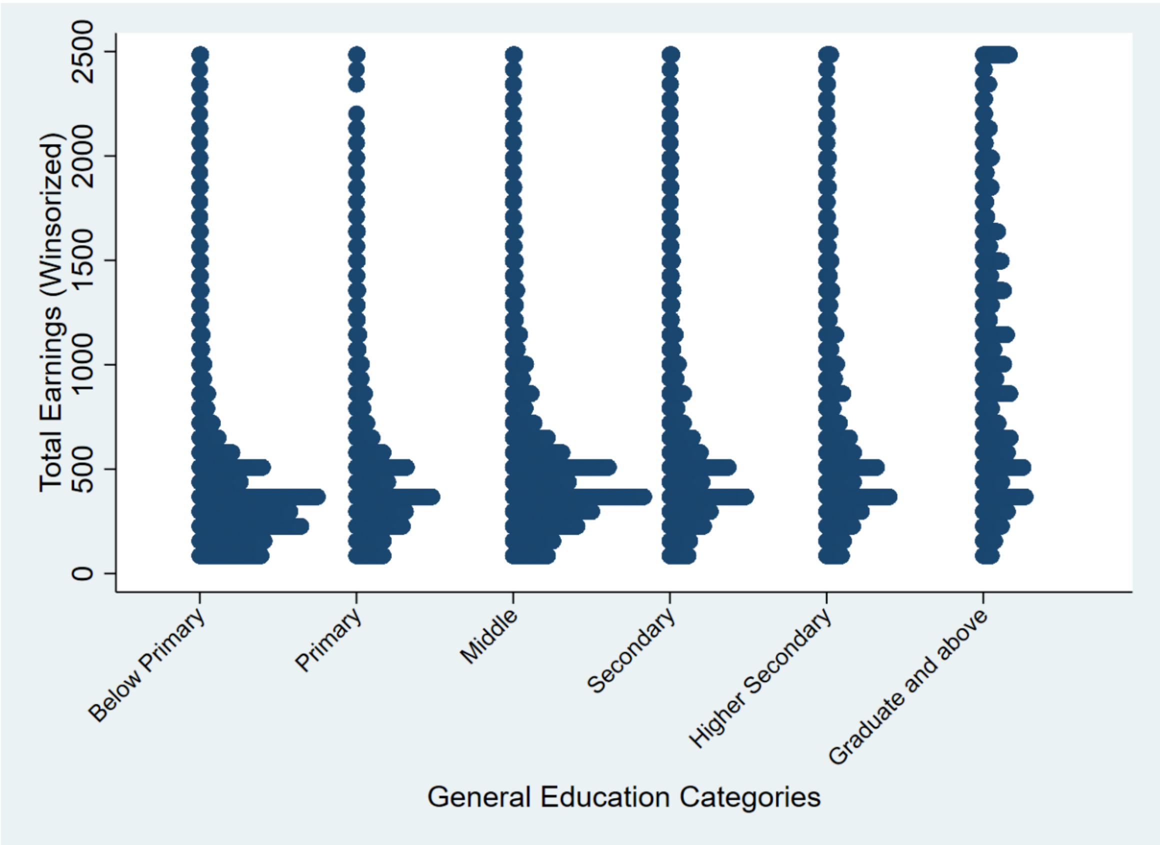
In contrast, formal training is more prevalent in knowledge-intensive sectors where specialized skills and certifications are requisite. The highest proportions of formal training participation are observed in Education (22%), Health & Social Work (20%), Financial & Insurance Activities (20%), and Information & Communication (20%), indicating that structured learning plays a pivotal role in these fields. These sectors typically necessitate professional qualifications, licenses, and regulatory compliance, rendering formal training an essential component of workforce preparation. However, despite the presence of formal training in certain industries, a substantial proportion of workers across multiple sectors remain without any training, highlighting potential skill deficiencies in the workforce.

Figure 4.3: Distribution of log of daily earnings by VET status



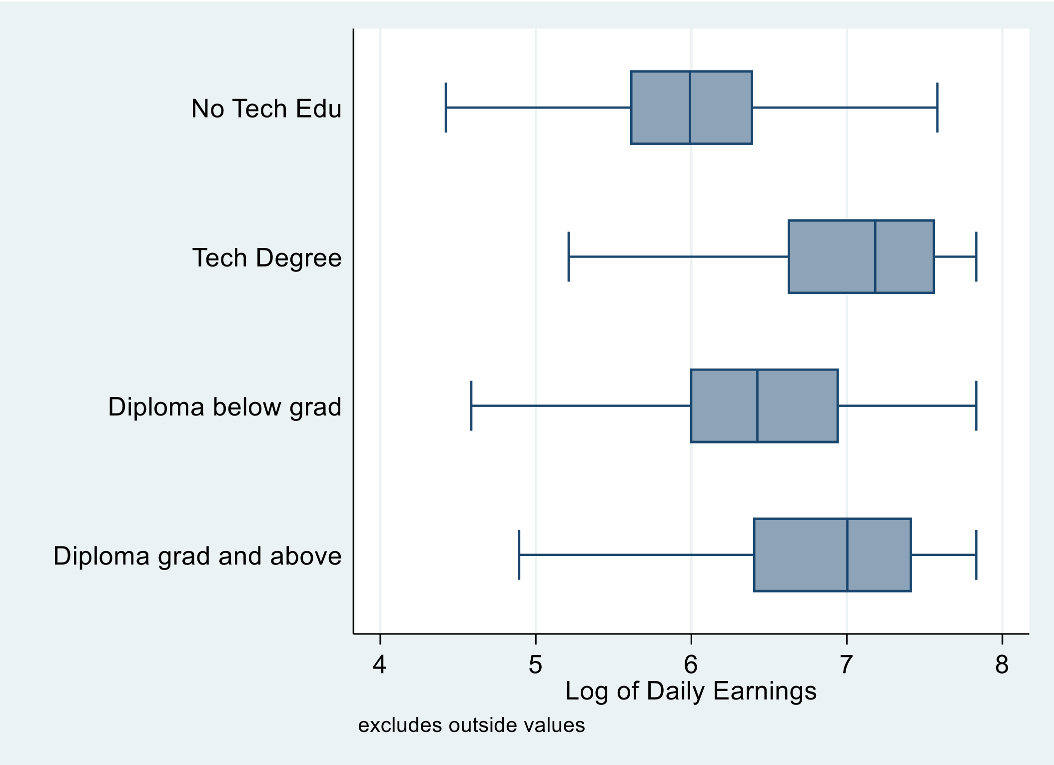
The diagram shows the distribution of daily earnings for each employment category depending on the training type they received. The dispersion in daily earnings is much higher among the regular wage employed compared to the self-employed. The dispersion in earnings is least among the casual wage employed. The median earnings are slightly higher for self-employed with formal training compared to the self-employed with informal training and no training. The median earnings for self-employed with informal training and no training don't have much difference. This implies that for the self-employed, informal training doesn't seem to yield a premium over no training. Further, the formal training might give a premium for the self-employed, but it is marginal. The regular or salaried employees secure a higher premium for formal training in terms of daily earnings compared to informal training and no training scenarios. The informal training doesn't seem to provide a significant wage increase for regular or salaried employees compared to no training scenario.

Figure 4.4: Distribution of log of daily earnings by general education categories



The dot plot illustrates the distribution of total earnings (winsorized) across various general education categories, demonstrating a distinct correlation between educational attainment and income levels. Individuals with higher levels of education exhibit a broader range of earnings, indicating greater income variability. The distribution for those with Below Primary, Primary, and Middle education appears more compressed, with the majority of individuals earning relatively low wages and fewer outliers in higher income brackets. As education levels increase to Secondary and Higher Secondary, earnings become slightly more dispersed, suggesting enhanced earning potential. The most pronounced difference is observed in the Graduate and Above category, where earnings display the highest spread, including significantly higher values at the upper end. This trend substantiates the positive correlation between education and earning potential, underscoring the role of higher education in facilitating access to more lucrative opportunities.

Figure 4.5: Distribution of log of daily earnings by technical education categories



The box plot highlights the effect of technical qualifications on income by showing the distribution of total daily earnings (winsorized) across various technical education levels. The lowest earnings, with a comparatively small interquartile range (IQR) and little fluctuation, are earned by those without technical education. With a much wider IQR and a higher median, those with technical degrees earn the most, suggesting that these credentials have significant financial advantages. In comparison to graduates with diplomas, who face both higher median wages and greater variability, those with diplomas below graduation level earn more than those without technical education but have a smaller earnings spread. According to the data, technical education—especially at the degree level—is linked to higher earning potential and more income variability, which may be a reflection of the workforce's skill-based wage disparities and variety of career options.

4.2 Estimation results

4.2.1 OLS Regression results

The table below gives the summary statistics of the sample for the OLS regression model. It provides an overview of the sample's demographic, educational, and social characteristics

Table 4.2: Summary statistics of the sample

Variable	Mean/ Proportion
Formal VET	4.78
Informal VET	30.68
No VET	64.54
Years of formal education	8.937
Age	34.189
Gender	
Male	50.18
Female	49.82
Religion	
Hindu	73.85
Muslim	14.96
Christian	7.04
Others	4.15
Sector	
Rural	56.57
Social Group	
Scheduled Tribes	14.61
Scheduled Castes	17.57
Other Backward Classes	41.27
Others	26.54
Marital Status	
Not currently married	34.63
Currently married	65.37
Region dummy	
North	28.35
South	19.86
East	30.09
West	21.70

A majority of individuals (64.54%) have not received any vocational education and training (VET), while only 4.78% have undergone formal VET, indicating limited access to structured vocational training. The average education level is approximately 8.94 years, suggesting that most individuals have completed middle or early secondary schooling. The sample is almost evenly split between males (50.18%) and females (49.82%), with an average age of 34.19 years, indicating a mix of young and mid-career individuals. In terms of social composition, the sample is predominantly Hindu (73.85%), with Muslims (14.96%), Christians (7.04%), and others (4.15%) making up the rest. More than half (56.57%) of the respondents reside in rural areas, highlighting the importance of rural employment and training dynamics. The representation of marginalized social groups is notable, with Scheduled Tribes (14.61%), Scheduled Castes (17.57%), and Other Backward Classes (41.27%) forming a significant portion of the sample. Additionally, 65.37% of individuals are currently married. Regionally, the East (30.09%) has the highest representation, followed by the North (28.35%), West (21.70%), and South (19.86%).

4.2.1.1 Formal versus Informal Vocational Education and Training

The returns to vocational education and training are estimated for the age group of 15-59 years. The marginal effects of vocational education after controlling for human capital variables (education, age squared as a proxy for experience), individual characteristics (age, gender, marital status), household characteristics (religion, social group, sector), and regional effects are displayed in Table 4.3.

Comparing the models, the Adjusted R-squared increases from 0.027 in Model 1 to 0.39 in Model 3, indicating that Model 3 explains the most variance. The univariate log-linear model indicates that formal vocational education and training (VET) leads to a 44.33% increase in daily earnings compared to those who have received informal VET. When controlling for human capital variables—such as years of formal education and age—as well as personal and household characteristics like gender, religion, sector, social group, marital status, and region, the estimated coefficient for formal VET experiences a significant reduction, declining by 27 percentage points. Additionally, when a non-linear regressor of age squared (which serves as a proxy for experience) is included in the model, it shows that formal VET results in 16.65% higher daily earnings compared to informal VET. Importantly, the coefficient for the formal VET dummy remains significant across all three models.

The coefficients of all explanatory variables remain significant in all three models. The number of years of formal education positively impacts daily earnings, with each additional year resulting in an average increase in daily earnings by 3.69% in Model 2 and 3.66 % in Model 3. The coefficient of age is positive and significant in Model 2 and Model 3 but the coefficient of age squared term is negative and significant in Model 3. This implies that daily earnings increase with one additional year of age (a proxy for additional years of experience), but this effect is non-linear and this is evident from the negative and significant coefficient of age squared term. This implies that earnings increase with age (a proxy for experience), but the rate of increase slows down as people get older. The positive and significant coefficient of the gender dummy for males implied that males earn 95.2% higher daily earnings on average compared to females according to Model 3. The coefficient of dummy variables for religious groups indicates that Muslims earn 1.5% lower daily compared to Hindus, Christians earn 15.25% higher daily compared to Hindus, and other religious groups clubbed together earn 10.6% higher compared to Hindus.

The sector dummy coefficient reveals that urban residents earn 26.3% more than those in rural areas (Model 3). This emphasizes the higher earnings typically associated with urban employment. The social groups of Scheduled Tribes, Scheduled Castes, and Other Backward Classes earn less than the "others" category. Specifically, compared to the "others," Scheduled Tribes have daily earnings that are 12% lower, Scheduled Castes earn 11.7% less, and Other Backward Classes earn 10.7 % less. In Model 2, being married is associated with a 5.5% increase in earnings. However, in Model 3, when age squared is included, this effect becomes much lower and its sign turns negative (-0.6%). The impact of regions on earnings is represented by dummy variables for various areas, with the northern region serving as the baseline. Southern states have earnings that are 30.5% higher, while Western states earn 1.3% more than Northern states. In contrast, Eastern states have earnings that are 12.6% lower than those in Northern states.

Table 4.3: OLS Regression results- Formal versus Informal VET

Explanatory Variables	**Model 1**		**Model 2**		**Model 3**	
	Coefficient	SE	Coefficient	SE	Coefficient	SE
Informal VET						
Formal VET	**0.367*** **	(0.020)	**0.159*** **	(0.0179)	**0.154*** **	(0.0177)
Years of formal education			**0.0369*** **	(0.00108)	**0.0366*** **	(0.00107)
Age			**0.00927*** **	(0.00041)	**0.0466*** **	(0.0028)
Gender						
Female (Ref)						
Male			**0.662*** **	(0.0119)	**0.669*** **	(0.0119)
Religion						
Hindu (Ref)						
Muslim			**-0.0217*** **	(0.0122)	**-0.0160*** **	(0.0121)
Christian			**0.145*** **	(0.0249)	**0.142*** **	(0.0251)
Other			**0.103*** **	(0.0267)	**0.101*** **	(0.0266)
Sector						
Rural (Ref)						
Urban			**0.237*** **	(0.0104)	**0.234*** **	(0.0103)
Social Group						
Others (Ref)						
Scheduled Tribe (ST)			**-0.132*** **	(0.0177)	**-0.128*** **	(0.0176)
Scheduled Caste (SC)			**-0.127*** **	(0.0131)	**-0.125*** **	(0.0131)
Other Backward Class (OBC)			**-0.116*** **	(0.0113)	**-0.114*** **	(0.0113)
Marital Status						
Not Married (Ref)						
Married			**0.0538*** **	(0.0104)	**-0.00615*** **	(0.0113)
Region						
North (Ref)						
South			**0.271*** **	(0.0137)	**0.266*** **	(0.0136)
East			**-0.134*** **	(0.0135)	**-0.135*** **	(0.0133)
West			**0.0118*** **	(0.0143)	**0.0127*** **	(0.0142)
Age Squared					**-0.000474*** **	(0.0000352)
Constant	**6.011*** **	(0.00693)	**4.812*** **	(0.025)	**4.175*** **	(0.0536)
Observations (N)	189669026		189665400		189665400	
Adjusted R^2	0.023		0.392		0.397	

4.2.1.2 Formal versus No Vocational Education and Training

Returns to formal VET compared to no VET for the age group of 15-59 are estimated using the OLS model, as done in the previous section.

Table 4.4: OLS Regression results- Formal versus No VET

Explanatory Variables	Model 1		Model 2		Model 3	
	Coefficient	SE	Coefficient	SE	Coefficient	SE
No VET						
Formal VET	**0.382*** **	(0.0201)	**0.133*** **	(0.019)	**0.130*** **	(0.0189)
Years of formal education			**0.0407*** **	(0.00118)	**0.0402*** **	(0.00118)
Age			**0.00894*** **	(0.000471)	**0.0385*** **	(0.00305)
Gender						
Female (Ref)						
Male			**0.552*** **	(0.0106)	**0.560*** **	(0.0108)
Religion						
Hindu (Ref)						
Muslim			**-0.0265*** **	(0.0164)	**-0.0235*** **	(0.0164)
Christian			**0.109*** **	(0.0243)	**0.108*** **	(0.0244)
Other			**0.0786*** **	(0.0239)	**0.0743*** **	(0.0238)
Sector						
Rural (Ref)						
Urban			**0.261*** **	(0.0116)	**0.259*** **	(0.0117)
Social Group						
Others (Ref)						
Scheduled Tribe (ST)			**-0.0704*** **	(0.0201)	**-0.0680*** **	(0.0202)
Scheduled Caste (SC)			**-0.128*** **	(0.0156)	**-0.128*** **	(0.0158)
Other Backward Class (OBC)			**-0.130*** **	(0.0136)	**-0.130*** **	(0.0136)
Marital Status						
Not Married (Ref)						
Married			**0.0876*** **	(0.0107)	**0.0433*** **	(0.0119)
Region						
North (Ref)						
South			**0.180*** **	(0.013)	**0.178*** **	(0.013)
East			**-0.139*** **	(0.0152)	**-0.140*** **	(0.0152)
West			**-0.00601*** **	(0.0149)	**-0.00413*** **	(0.0149)
Age Squared					**-0.000376*** **	(0.0000378)
Constant	**5.996*** **	(0.00739)	**4.865*** **	(0.0267)	**4.361*** **	(0.06)
Observations (N)	194529421		194529421		194529421	
Adjusted R^2	0.023		0.368		0.371	

The results from three log-linear models estimating the returns to formal vocational education and training vis-à-vis those who didn't receive any training are shown in Table 4.4. The adjusted

R-squared value is highest in model 3 which means it explains higher variation compared to the other two models. The coefficient of the formal VET dummy in three models shows that the magnitude of the coefficient gets smaller as controls with statistically significant coefficients are added to the model. Model 1 shows that formal VET yields 46.5% higher earnings compared to those who have no VET. But these earnings premium declines to 14.2% in model 2 and 13.8% in model 3. All the other controls included in the models have statistically significant coefficients. One additional year of formal education yields 4.02% higher daily earnings after controlling for other covariates. The age has a positive and significant coefficient but the variable age squared has a negative and significant coefficient. This implies that earnings increase with age (a proxy for experience), but the rate of increase slows down as people get older. The gender dummy shows that males earn more than females. The sign of coefficients of sector dummy and religious category dummies are the same as those in the previous model. The currently married have higher earnings than those who are currently not married. The Southern states have higher daily earnings on average compared to northern states, but eastern and western states have lower earnings compared to northern states.

4.2.1.3 Informal versus No Vocational Education and Training

Table 4 shows the results from three log-linear models estimating the returns to informal vocational education and training vis-à-vis those who didn't receive any training. The adjusted R-squared value is highest in model 3 which means it explains higher variation compared to the other two models. The coefficient of the informal VET dummy in the models shows that the magnitude of the coefficient gets smaller as controls with statistically significant coefficients are added to the model. Model 1 shows that informal VET yields 1.5% higher earnings compared to those who have no VET. But these earnings premium declines to 0.6% in model 2 and 0.5% in model 3. This implies that returns to informal VET are lower than formal VET when compared to the category of no VET. The coefficients of other controls have similar signs as those of previous OLS models. The years of formal education have a positive and significant coefficient. Experience (proxied by age) yields positive but diminishing returns.

Table 4.5: OLS Regression results- Informal versus No VET

Explanatory Variables	Model 1		Model 2		Model 3	
	Coefficient	SE	Coefficient	SE	Coefficient	SE
No VET						
Informal VET	**0.0150*****	(0.00968)	**0.00654*****	(0.00701)	**0.00513*****	(0.00699)
Years of formal education			**0.0365*****	(0.00085)	**0.0362*****	(0.000847)
Age			**0.00799*****	(0.00033)	**0.0412*****	(0.00215)
Gender						
Female (Ref)						
Male			**0.611*****	(0.00866)	**0.620*****	(0.00873)
Religion						
Hindu (Ref)						
Muslim			**-0.0178*****	(0.0108)	**-0.0132*****	(0.0108)
Christian			**0.116*****	(0.0179)	**0.115*****	(0.0179)
Other			**0.111*****	(0.0182)	**0.107*****	(0.0181)
Sector						
Rural (Ref)						
Urban			**0.238*****	(0.00846)	**0.236*****	(0.00847)
Social Group						
Others (Ref)						
Scheduled Tribe (ST)			**-0.122*****	(0.0145)	**-0.118*****	(0.0144)
Scheduled Caste (SC)			**-0.128*****	(0.0112)	**-0.126*****	(0.0112)
Other Backward Class (OBC)			**-0.126*****	(0.00965)	**-0.125*****	(0.00966)
Marital Status						
Not Married (Ref)						
Married			**0.0627*****	(0.00778)	**0.0103*****	(0.00873)
Region						
North (Ref)						
South			**0.238*****	(0.0101)	**0.234*****	(0.0101)
East			**-0.135*****	(0.0106)	**-0.136*****	(0.0106)
West			**0.0122*****	(0.0112)	**0.0136*****	(0.0113)
Age Squared					**-0.000421*****	(0.0000267)
Constant	**5.996*****	(0.00739)	**4.900*****	(0.0196)	**4.332*****	(0.0426)
Observations (N)	348516155		348512529		348512529	
Adjusted R^2	0.000		0.373		0.377	

4.2.2 Propensity Score Matching (PSM)

The study employs propensity score matching to analyze the differing impacts of various types of vocational training on daily earnings. It focuses on three outcome variables: overall daily earnings (without distinguishing between self-employed, regular or salaried employed, and casual labour), daily earnings of the self-employed, and daily earnings of regular or salaried employees.

Additionally, the research seeks to understand how different types of formal training—namely on-the-job, part-time, and full-time training—affect the outcome variables, including overall daily earnings and the daily earnings of regular or salaried employees.

4.2.2.1 Impact of types of VET on earnings

4.2.2.1.1 Formal versus Informal VET

The impact of formal VET on daily earnings compared to informal VET is analyzed using the propensity score matching approach. This study uses the caliper matching approach to match the treated and control units. The set of covariates is used in the logit model to estimate the propensity score of the observations in the treatment group and control group. The full estimates of the logit model are provided in Table A2 in the appendix. The p-value (0.000) of the Likelihood Ratio (LR) chi-square test indicates that in the logit model, at least one independent variable significantly predicts the probability of receiving formal VET vis-à-vis informal VET. Furthermore, the pseudo-R-squared value of 0.2296 (22.96%) indicates that the model accounts for approximately 22.96% of the variation in the dependent variable. In the logit model, all covariates are significant except for the social group dummy representing scheduled castes and the religious group dummy representing Others (those belonging to religious groups other than Hindu)

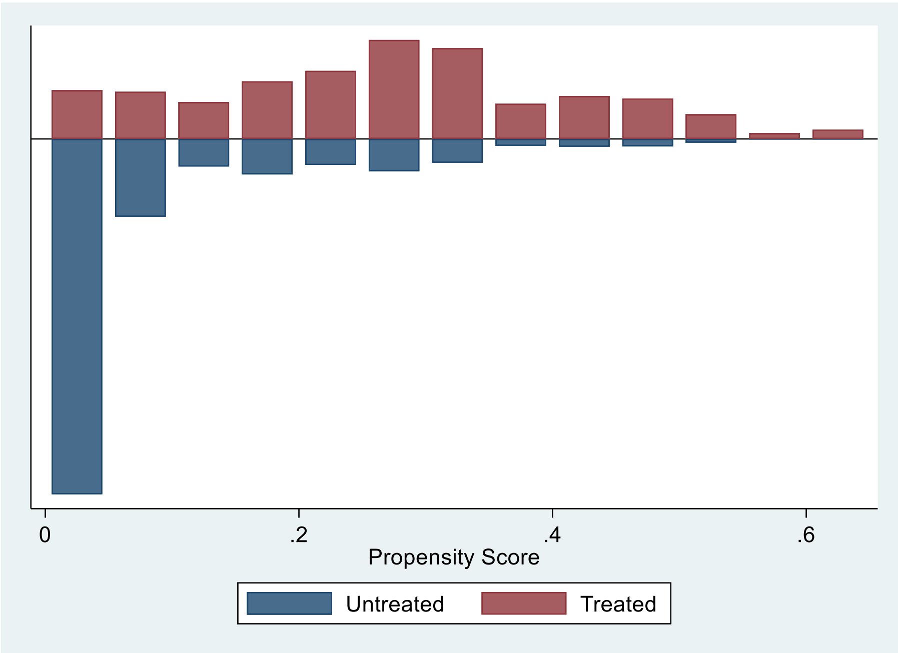
Table 4.6: Propensity Score Matching Results: Average Treatment Effect on the Treated (Formal versus Informal VET)

	Treated (Formal VET)	Controls (Informal VET)	Difference	SE	T-stat
Unmatched	6.430	6.052	0.378	0.009	44.23
ATT	6.430	6.300	0.130	0.028	4.65

Table 4. indicates that before matching, there was a substantial difference (T = 44.23) between the treated and control groups. This suggests that the two groups were not well balanced before matching, which raises concerns about potential selection bias affecting the results. Therefore, this indicates that the OLS (Ordinary Least Squares) results may be biased. After matching, the difference between the treated and control groups has decreased from 0.378 to 0.130 implying that the treatment effect is much lower. The t statistic of 4.65 suggests that the difference is still

significant. The results show that the treatment (formal VET) increases total daily earnings by 13% compared to the matched control group (Informal VET).

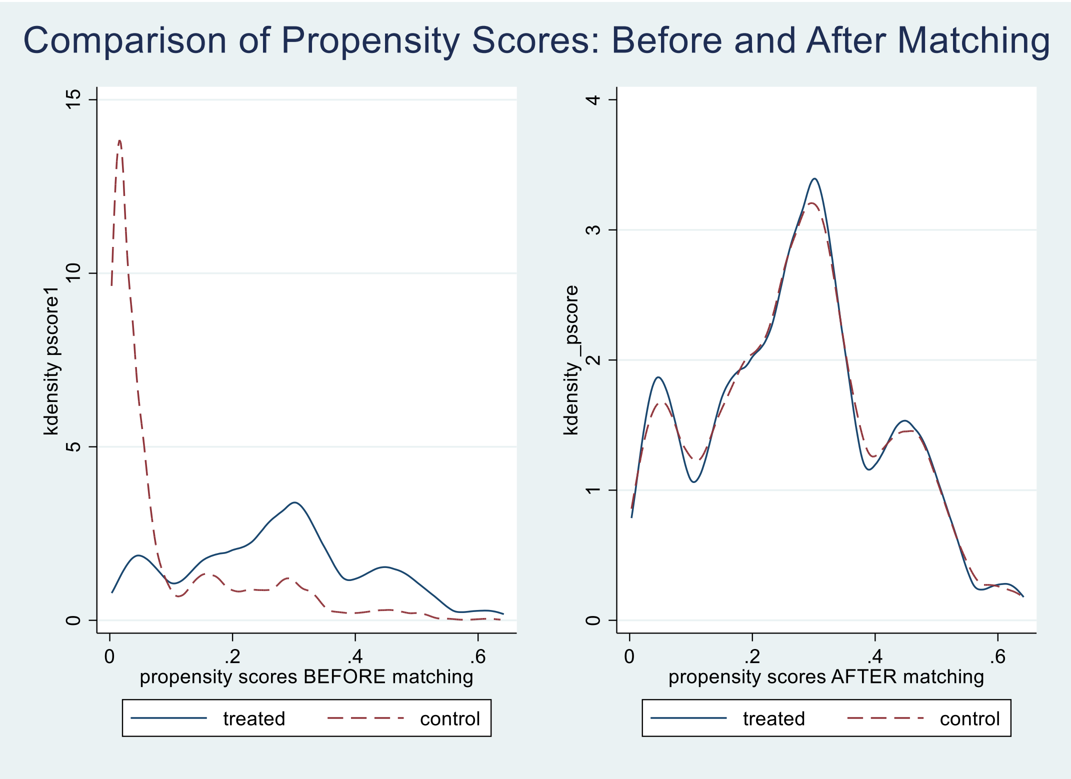
Figure 4.6: Propensity score distributions of treated and control groups (Formal versus Informal VET)- Outcome variable: Overall daily earnings



In propensity score matching, common support is crucial because it guarantees that the features of the treated and control units are similar, enabling a reliable assessment of treatment effects. The above diagram shows that though there is a lack of overlap at the high propensity scores, there is better overlap at low to mid propensity scores. The graphical evidence points towards ample common support.

To ensure that the propensity score matching results are valid, there is a need to check the balance of covariates between treated and control groups, before and after matching. A high pseudo-R2 (0.231) before matching indicates significant differences between the treated and control groups. Following matching, pseudo-R2 falls to 0, indicating that there is no statistical difference between the treated and control groups in terms of covariates. This attests to effective balancing. There is selection bias since the Likelihood Ratio test is very significant ($p < 0.05$) before matching, indicating that variables substantially influence treatment assignment. Covariates no longer predict treatment assignment after matching, as indicated by $p = 0.989$, confirming balance.

Figure 4. 7: Comparison of propensity score before and after matching- Formal versus Informal VET (Overall daily earnings)



The panel on the left shows the distribution of propensity scores for the treated (solid blue line) and control (dashed red line) groups is substantially different before matching. The treated group is more equally distributed, but the control group has a much larger density at lower propensity scores. This implies that before matching, the treated and control groups' observed attributes were out of balance.

The panel on the right shows that the treated and control groups' propensity score distributions start to resemble each other more after matching. This shows that the two groups' propensity scores have been successfully balanced by the matching procedure. The distributions' closeness indicates that selection bias is successfully reduced by the matching procedure, enhancing comparability.

<u>4.2.2.1.2 Formal versus no VET</u>

The propensity score matching method is used to create a counterfactual and assess the impact of formal VET on daily earnings compared to no VET. The p-value (0.000) of the Likelihood Ratio (LR) chi-square test indicates that in the logit model, at least one independent variable significantly predicts the probability of receiving formal VET vis-à-vis no VET. Furthermore, the pseudo-R-squared value of 0.2075 (20.75%) indicates that the model accounts for approximately 20.75% of the variation in the dependent variable. In the logit model, coefficients of the social group dummy representing scheduled castes and the social group dummy for obc are not significant.

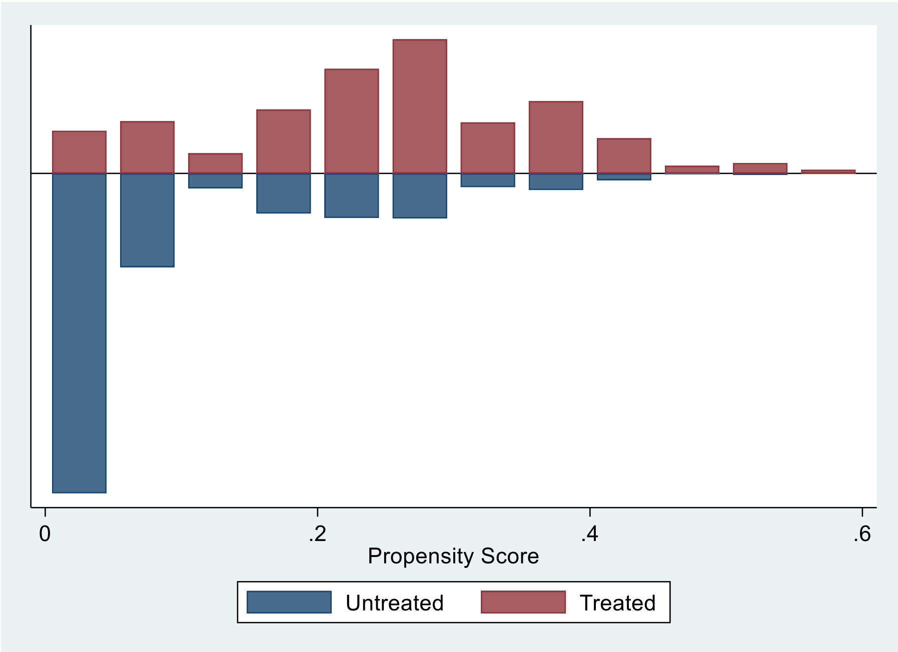
Table 4.7: Propensity Score Matching Results: Average Treatment Effect on the Treated (Formal versus No VET)

	Treated (Formal VET)	Controls (No VET)	Difference	SE	T-stat
Unmatched	6.430	6.050	0.381	0.009	44.03
ATT	6.430	6.318	0.112	0.027	4.10

The differences between the treated and control groups' log-transformed total earnings before and after propensity score matching are shown in Table 4.7.

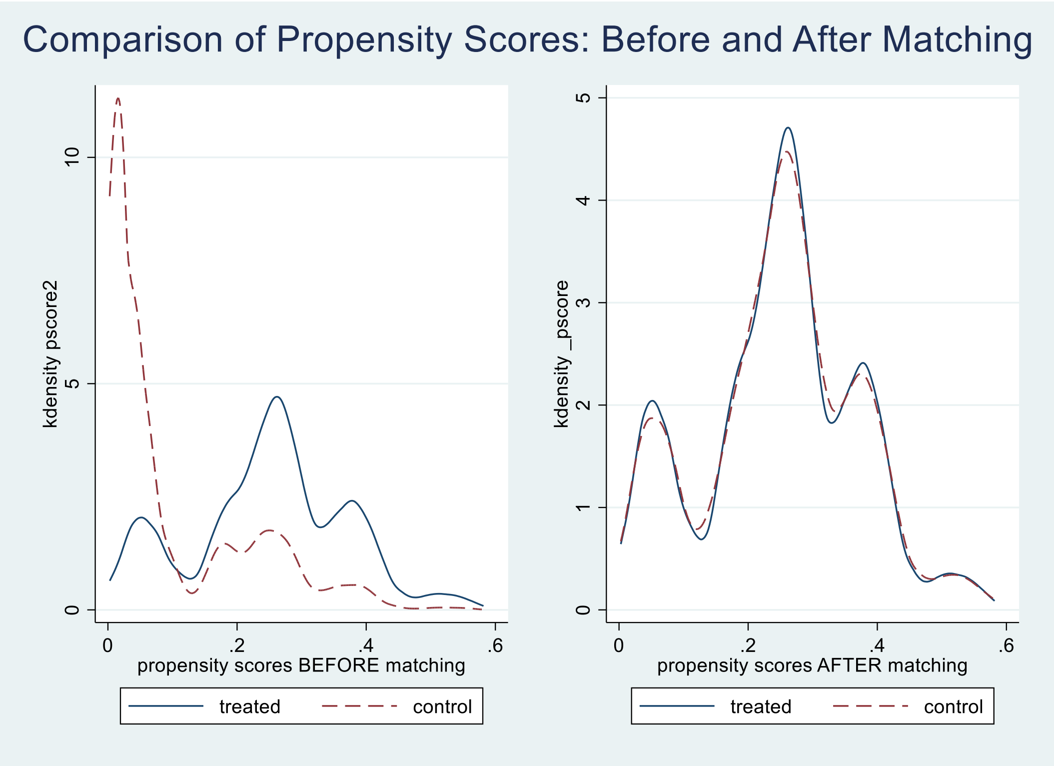
Before matching (unmatched row), the treated group's average log earnings were 6.430, while that of the control group was 6.050. Strong pre-treatment differences between the two groups are shown by the statistically significant difference of 0.381, which has a T-statistic of 44.03 and a very small standard error of 0.009. This implies that the treatment and control groups' wages were not comparable before matching. With a higher standard error (0.027) and a T-statistic of 4.1 following matching (ATT row), the difference in log earnings between the treated and matched control groups drops to 0.112, which is statistically significant. This indicates that there is evidence showing that treatment (formal VET compared to no VET) significantly affects earnings, resulting in an 11.2% earnings premium, even after accounting for observable differences using propensity score matching. By balancing the variables and lowering selection bias, matching has enhanced comparability between the groups, as evidenced by the narrowing of the earnings disparity.

Figure 4.8: Propensity score distributions of treated and control groups (Formal versus No VET)- Outcome variable: Overall daily earnings



The distribution of propensity scores for individuals who received formal Vocational Education and Training (VET) (treated group) and those who did not (control group) are shown in the diagram. At the lowest end of the propensity score distribution, specifically between 0.0 and 0.1, where it is most prevalent, is the control group (no VET). The treated group (formal VET), on the other hand, is more uniformly distributed throughout the mid-range of propensity scores, with its peak concentration situated between 0.2 and 0.3. Notably, when there are fewer treated people present, the control group predominates at extremely low propensity scores (around 0.0). On the other hand, both groups exhibit overlapping distributions for propensity scores between around 0.1 and 0.5, which is important for identifying similar people across treatment conditions. Both categories become less prevalent beyond 0.5, indicating fewer observations with high estimated probabilities of obtaining VET.

Figure 4. 9: Comparison of propensity score before and after matching- Formal versus No VET (Overall Daily-earnings)



The distribution of propensity scores for the treatment and control groups prior to and following matching is contrasted in the figure. The density of propensity scores prior to matching is depicted in the left panel, which demonstrates a notable disparity between the two groups. The treatment group (solid blue line) is more uniformly distributed than the control group (dashed red line), which has a greater concentration of observations at lower propensity scores. This implies that the treated and control groups were not comparable with regard to their chances of receiving treatment prior to matching, which may introduce bias into the estimation of treatment effects.

The density of propensity scores upon matching is shown in the right panel, showing a significantly better alignment between the treated and control groups. The fact that the two distributions now more widely overlap suggests that the matching procedure was successful in balancing the two groups' observable traits. This better balance implies that selection bias has been successfully mitigated by propensity score matching (PSM), increasing the comparability of the treated and control groups to draw conclusions about causality. Therefore, any post-matching disparities in the two groups' outcomes are more likely to be due to the treatment than to underlying differences in the groups' characteristics.

4.2.2.1.3 Informal versus No VET

The propensity score matching was conducted to estimate the impact of informal vocational education and training (VET) compared to no VET on daily earnings, in a manner similar to the methods explained in the previous sections. The logistic regression results show that the model is significant with a low pseudo-R-squared value of 0.0109.

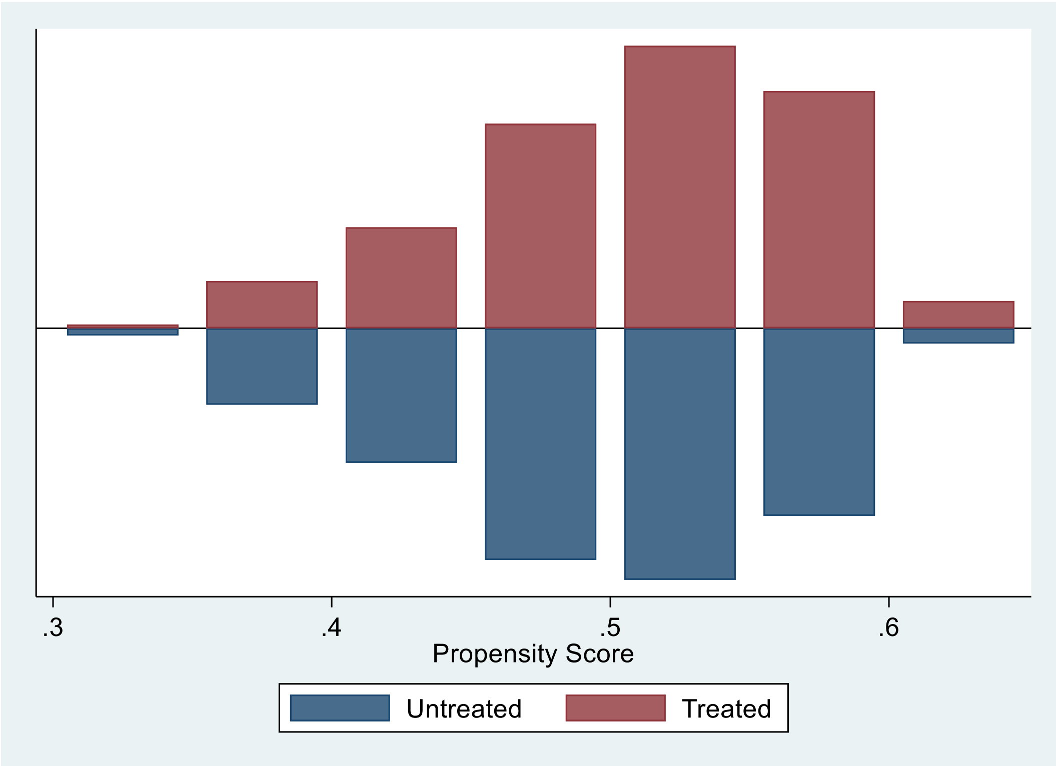
Table 4.8: Propensity Score Matching Results: Average Treatment Effect on the Treated (Informal versus No VET)

	Treated (Informal VET)	Controls (No VET)	Difference	SE	T-stat
Unmatched	6.052	6.050	0.003	0.004	0.65
ATT	6.052	6.069	-0.017	0.016	1.06

By contrasting treated people (who received informal VET) with controls (who received no VET) before and after propensity score matching, Table 4.8 illustrates the effect of informal VET on the log of daily earnings. The treatment group's average log of daily earnings before matching (Unmatched row) is 6.052, whereas that of the control group is 6.050. With a standard error of 0.004 and a t-statistic of 0.65, the 0.002 difference is negligible and not statistically significant.

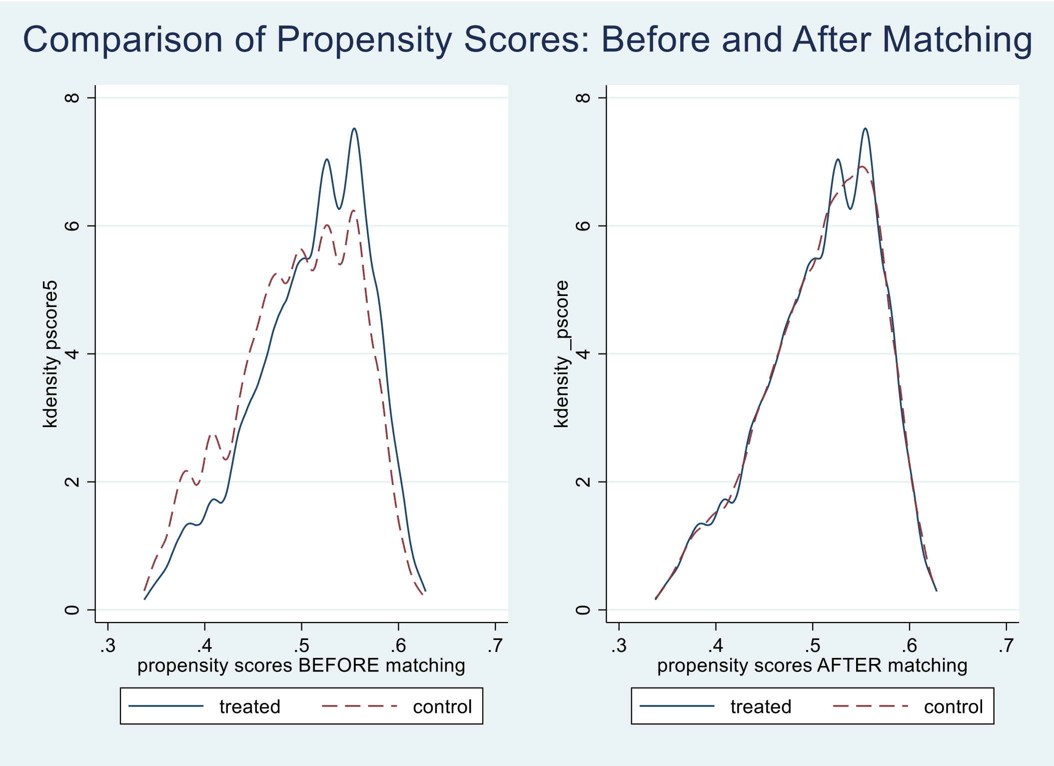
Following matching, the treated group's average log of daily earnings stays at 6.052, while the matched control group's log earnings marginally rise to 6.069. Those who got informal VET have, on average, somewhat lower log daily wages than those who did not acquire VET, according to the estimated treatment impact (ATT), which is currently -0.017. With a t-statistic of 1.06 and a standard error of 0.016, this difference is statistically insignificant. This implies that, at least for this sample, informal VET has no discernible effect on income. Because the matching procedure effectively balanced the groups' visible features, the reported earnings gap is unlikely to be the result of pre-existing differences and instead represents the actual lack of a significant earnings difference.

Figure 4.10: Propensity score distributions of treated and control groups (Informal versus No VET)- Outcome variable: Overall daily earnings



The histogram of propensity scores for treated and untreated groups is depicted in the picture. Given observed covariates, the propensity score—which is represented by the x-axis—is the likelihood of receiving treatment. The y-axis shows the frequency of observations; the treated group is represented by bars above the axis (in red), and the untreated group is represented by bars below the axis (in blue). How well the two groups are balanced in terms of observed characteristics is indicated by the overlap of the treated and untreated groups across propensity score ranges. The matching process appears to have found similar people in both groups if there is a substantial overlap between them, especially between 0.4 and 0.6. Less overlap exists in some regions, particularly at the lower and higher ends of the propensity score range, suggesting that some members of the treated or untreated groups do not have counterparts with comparable scores.

Figure 4. 11: Comparison of propensity score before and after matching- Informal versus No VET (Overall Daily-earnings)



The kernel density estimation of propensity scores for the treated and control groups before matching is displayed in the left panel. Given that the treated group has a wider spread and the control group has a higher density at lower propensity scores, the apparent discrepancy in distributions suggests an imbalance. Causal inferences may be biased due to this imbalance, which indicates that the two groups differ significantly in terms of observed covariates before matching. The kernel density estimates following matching are shown in the right panel, demonstrating how the propensity score distributions of the treated and control groups have significantly improved. The fact that the two lines nearly overlap shows that the matching procedure was successful in enhancing group balance. By lowering selection bias, this alignment improves the comparability of the treated and control groups when estimating the effects of treatment.

4.2.2.2 Impact of types VET on earnings of self-employed and earnings of regular or salaried employees

The impact of different types of vocational education and training on the daily earnings masks the heterogenous impacts these VET types might have on the earnings of the self-employed as well as on the regular or salaried employees. So, it is necessary to do the analysis separately for the self-employed and the regular or salaried employees.

Table 4.9: Propensity Score Matching Results: Average Treatment Effect on the Treated

Outcome: Log Daily Earnings		Formal vs Informal VET	Formal vs No VET	Informal vs No VET
Regular/Salaried Employed	ATT	0.112	0.110	0.021
	T-Stat	3.74	3.63	1.07
Self-Employed	ATT	0.063	0.035	-0.029
	T-Stat	1.96	1.06	-1.47

The table 4.9 shows the Average Treatment Effect on the Treated (ATT) as well as the t-statistic associated with different types of treatment (types of VET) on the outcome variables of daily earnings of self-employed as well as daily earnings of the wage/salaried employed.

Formal VET appears to have a strong positive effect on earnings for individuals in regular or salaried employment. Those who completed formal VET earned 11.2 percentage points more than those who completed informal VET, with a statistically significant t-statistic of 3.74. Similarly, those with formal VET are 11 percent more likely to earn more than those without VET, with a t-statistic of 3.63, again indicating strong statistical significance. These findings suggest that formal VET provides meaningful skills upgrading leading to higher earnings in structured employment. On the other hand, informal VET provides marginal gains in earnings, with the ATT estimate of 0.021 (or 2.1) being not statistically significant (t-stat = 1.07). This suggests that informal VET, which can include hereditary training or self-learning, does not translate into significant wage increases in the formal employment environment. Employers may perceive formal VET qualifications as more reliable, which will lead to better job matching and higher pay for those with formal VET qualifications.

Among self-employed workers, the effects of VET are more subtle. Compared to informal VET, formal VET leads to an increase in earnings of 6.3 percent, t-statistic of 1.96 indicating marginal relevance at 5 percent. However, when comparing formal VET with non-formal VET, the increase in earnings is much smaller (3.5%) and statistically insignificant (t-stat = 1.06). Surprisingly, informal VET learners seem to earn 2.9 percent less than those without VET, with a negative ATT of -0.029 and t-statistic of -1.47, suggesting that having informal VET doesn't help much in raising the earnings of self-employed. Although not statistically significant, this result suggests that informal VET may not provide self-employed workers with the skills they need to increase their income effectively. In some cases, informal training may focus on narrow skills which are not well suited to profitable self-employment.

The figures A1 to A6 in the appendix shows that there was common support for the treatment group. Further, the figures A7 to A12 in the appendix show the propensity scores for treatment groups before and after matching which shows that there was proper matching. The results of the Likelihood Ratio (LR) Chi-squared test, which assesses whether the covariates jointly explain treatment assignment, are shown in Table A8 in the appendix. It shows that the LR Chi-squared value has significantly reduced in all the models after matching which indicates successful matching. The LR Chi-squared value was significant only in the case of informal versus no VET models, indicating some imbalance between the treatment group and the matched control group, but the significant reduction in mean bias and median bias post-matching add validity to the results of propensity score matching.

5. Discussion

The results of this study are more or less in line with the findings of the previous studies. Many previous studies have found a positive significant impact of vocational education and training on wages in the Indian context. This study attempts to contribute to this stream of literature which tries to estimate the returns to vocational education and training in India.

Table 6.1: OLS Regression results by VET Category

OLS Regression Results for Log of Daily Earnings by VET Category

	Formal vs Informal VET	Formal vs No VET	Informal vs No VET
Log of daily earnings	16.65*** (0.0177)	13.8*** (0.0189)	0.5*** (0.00699)

Table 6.1 shows the results from ordinary least squares estimated in this study. Formal vocational education and training (VET) offers 16.65% higher earnings compared to informal VET and 13.8% higher returns than having no VET at all. In contrast, informal VET provides only a

minimal earnings premium of 0.5% compared to not participating in VET. These models don't address the selection bias arising from self-selection into the labour force. Further, they also don't address the endogeneity associated with the choice of vocational education and training.

Numerous studies in the Indian context have reported similar findings. Agarwal and Agarwal (2017) employed the Heckman method for selectivity correction and discovered that vocational education offers a 20% higher return compared to general secondary education. Kumar, Mandava, and Gopanappali (2019), using multiple linear regression, found that overall formal training increases wages by 4.7%. They also noted that returns on formal training are significant in the primary sector (36.9%) and the secondary sector (17.6%), but not in the tertiary sector.

Bahl, Bhatt, and Sharma (2021), in a study also utilizing the Heckman method, found similar outcomes. They identified a wage premium of 12.5% for formal vocational education and training (VET) compared to informal VET. Furthermore, the wage premium for formal VET compared to having no VET was 13.5%, while the premium for informal VET compared to no VET was 3%. Banerjee (2016) reported a wage premium of 28% associated with formal VET within the manufacturing sector.

All of these studies focused exclusively on wage earnings and did not consider the earnings of the self-employed. The present study includes the earnings of the self-employed in the analysis.

Table 6.2: Propensity Score Matching Results

Outcome		Formal vs Informal VET	Formal vs No VET	Informal vs No VET
Log daily earnings	ATT	0.13	0.112	-0.017
	T-Stat	4.65	4.1	1.06
Log earnings of Regular/Salaried employees	ATT	0.112	0.11	0.021
	T-Stat	3.74	3.63	1.07
Log earnings of Self-employed	ATT	0.063	0.035	-0.029
	T-Stat	1.96	1.06	-1.47

Table 6.2 summarizes the results from the propensity score matching exercise conducted in this study. It shows that formal vocational education and training (VET) provides an earnings premium over both informal VET and no VET when considering overall daily earnings, as well as the daily earnings of regular salaried employees. This difference in earnings is statistically significant. However, the average treatment effect on the treated (ATT) estimates indicates that informal VET does not significantly enhance earnings compared to having no VET, regarding both overall daily earnings and the daily earnings of regular or salaried employees.

When examining the earnings of self-employed individuals, formal VET also offers a premium over informal VET; however, this difference is only marginally significant. In terms of self-employed earnings, the different types of VET appear to have either a weak effect or no significant impact on earnings at all. The findings for the wage employed align with those of Bahl, Bhatt, and Sharma (2021), who noted that formal VET yields 15% higher returns compared to informal VET and 13% higher returns compared to having no VET. They also reported that the earnings difference between individuals with informal VET and no VET, following matching through propensity score matching (PSM), was minimal at 1.8%. In contrast, the present study found a negative difference that was not statistically significant. Importantly, Bahl, Bhatt, and Sharma (2021) did not include the earnings of self-employed individuals in their analysis.

6. Conclusion

There has been a heightened focus on skill development and vocational education in India with the launch of the Skill India Mission. This initiative aims to prepare youth for the future by equipping them with industry-relevant skills. The Government of India's Skill India Mission (SIM), administered by the Ministry of Skill Development and Entrepreneurship (MSDE), offers skill development, re-skilling, and up-skilling training through an extensive network of skill development centers and institutes.

In this context, it is essential to evaluate the returns on vocational education and training in India. This study seeks to estimate the impact of various types of vocational education and training on the daily earnings of different categories of employees, utilizing a nationally representative Periodic Labour Force Survey (PLFS) for the year 2023-24.

The survey data indicates that only 34.7% of individuals aged 12-59 have received some form of vocational education and training, with just 4.1% having received formal vocational education and training. This reveals that the coverage of formal vocational education and training in India is very low compared to informal forms of vocational education and training. The formally trained individuals seem to be employed in higher proportion in regular or salaried employment compared to those who received informal training. However, the unemployment rate also seems to be higher among the formally trained compared to informally trained as well as those who lacked any form of training. This might be because of the higher aspirations of the individuals who have received formal vocational education and training.

The results of the estimations indicate that, without accounting for selection bias associated with the choice to participate in the labour market and the issue of endogeneity, formal vocational education and training offers significantly higher returns compared to both informal vocational education and training, as well as no vocational education and training at all. Using the quasi-experimental method of propensity score matching, which addresses the issue of endogeneity related to the choice of vocational education and training, this study demonstrates that formal vocational education and training provides higher returns compared to both informal vocational education training and no vocational education at all. These higher returns are evident for overall daily earnings as well as for the daily earnings of regular salaried employees. However, this effect does not appear to be significant for self-employed individuals. Additionally, the earning premium associated with informal vocational education and training is not statistically significant when compared to no training.

The findings of the study indicate that formal training yields significant earnings premium. The different types of formal training are on-the-job training, part-time training and full-time training. Thus, employers should be incentivized to invest more in training, as this can lead to a higher earnings premium. Furthermore, there needs to be greater involvement from industry bodies and associations in shaping the training landscape. This approach will help ensure that training curricula are effectively aligned with the needs of the labour market and specific job requirements. By doing so, it may enhance the effectiveness of formal training programs offered in both part-time and full-time formats. Additionally, initiatives aimed at integrating individuals who have received informal training into formal programs should improve their earning potential and assist the country in reaping its demographic dividend.

References

1. Ahmed, T., & Chattopadhyay, R. (2016). Return to general education and vocational education & training in Indian context: Policy implications. International Journal of Educational Management, 30(3), 370-385.
2. Anikin, V. A. (2021). Training poverty in India: Reassessing the effect of social ascription. International Journal of Training and Development, 25(3), 277-300.
3. Aypay, A. (2003). The tough choice at high school door: an investigation of the factors that lead students to general or vocational schools. International Journal of Educational Development, 23(5), 517-527.
4. Banerjee, T. (2016). Impacts of vocational education and training on employment and wages in Indian manufacturing industries: variation across social groups—empirical evidences from the 68th round NSSO data. The Indian Journal of Labour Economics, 59, 489-509.
5. Becker, G. S. (1993). Nobel lecture: The economic way of looking at behavior. Journal of political economy, 101(3), 385-409.
6. Bennell, P., & Segerstrom, J. (1998). Vocational education and training in developing countries: Has the World Bank got it right? International Journal of Educational Development, 18(4), 271-287.
7. Billett, S. (1998). Enterprises and vocational education and training: expenditure and expected returns. Journal of Vocational education and Training, 50(3), 387-402.
8. Borghans, L., De Grip, A., & Heijke, H. (1996). Labor market information and the choice of vocational specialization. Economics of Education Review, 15(1), 59-74.
9. Carneiro, P., Hansen, K. T., & Heckman, J. J. (2003). Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College.
10. Cedefop (2021). Understanding technological change and skill needs: skills surveys and skills forecasting. Cedefop practical guide 1. Luxembourg: Publications Office.
11. Deaton, A. (1997). The analysis of household surveys: a microeconometric approach to development policy. World Bank Publications.
12. Debroy, B. (2009). Vocational Education and India's Skill Deficit. Institute of South Asian Studies.
13. Doerr, A., Fitzenberger, B., Kruppe, T., Paul, M., & Strittmatter, A. (2017). Employment and earnings effects of awarding training vouchers in Germany. ILR Review, 70(3), 767-812.

14. 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(1), 1-17.
15. Mincer, J. A. (1974). The human capital earnings function. In Schooling, experience, and earnings (pp. 83-96). NBER.
16. Moenjak, T., & Worswick, C. (2003). Vocational education in Thailand: a study of choice and returns. Economics of Education Review, 22(1), 99-107.
17. Moodie, G. (2002). Identifying vocational education and training. Journal of vocational education and training, 54(2), 249-266.
18. Psacharopoulos, G. (1987). To vocationalize or not to vocationalize? That is the curriculum question. International review of education, 33, 187-211.
19. Psacharopoulos, G., & Patrinos, H. A. (1993). Secondary vocational education and earnings in Latin America. The vocational aspect of education, 45(3), 229-238.
20. Psacharopoulos, G. (1995). The profitability of investment in education: concepts and methods.
21. Psacharopoulos, G. (2024). Returns to education: a brief history and an assessment. Education Economics, 32(5), 561-565.
22. Schultz Theodore, W. (1963). The economic value of education. New York.
23. Tilak, J. B. (2003). Vocational education and training in Asia. In International handbook of educational research in the Asia-Pacific region: Part one (pp. 673-686). Dordrecht: Springer Netherlands.
24. Tognatta, N. (2014). Technical and vocational education and training in India: A study of choice and returns. University of Pennsylvania.
25. Tunali, I. (2003). General versus. vocational secondary school choice and labor market outcomes in Turkey, 1988-98. In 10th annual ERF conference in Morocco.
26. Van Den Berg, N., Meijers, F., & Sprengers, M. (2006). More vocational education and supplementary training through equalization of costs? An analysis of a training and development fund in the Netherlands. Human Resource Development International, 9(1), 5-24.
27. Vincent, A., & Rajasekhar, D. (2023). Do different types of vocational education and training programmes influence earnings? Recent evidence from India. Education+ Training, 65(3), 454-469.

Appendix

Table A1: Form of Training by Employment Category and Gender

Training Type	Self Employed	Regular/ Salaried	Casual Labour	Unemployed	Not in Labour Force	Total
Male						
Formal	26.09	49.87	2.44	9.73	11.87	100
Hereditary	78.12	3.58	14.49	1.36	2.45	100
Self-learning	58.18	15.38	22.1	2.03	2.31	100
Learning on the Job	26.61	47.83	22.21	2.13	1.21	100
Others	54.44	24.27	14.86	2.58	3.86	100
No Training	25.92	14.87	14.28	4.83	40.09	100
Female						
Training Type	Self Employed	Regular/ Salaried	Casual Labour	Unemployed	Not in Labour Force	Total
Formal	20.44	23.59	0.75	8.64	46.59	100
Hereditary	60.82	1.28	7.46	0.53	29.91	100
Self-learning	53.43	5.34	6.13	1.33	33.76	100
Learning on the Job	32.47	40.18	10.91	1.81	14.62	100
Others	43.96	9.81	5.57	3.28	37.38	100
No Training	15.06	4.8	4.51	1.7	73.93	100

Table A2: Logit model (Choice of Formal versus Informal VET)

Variable	Coef.	Std. Err	z	P>z	[95% Conf. Interval]
General Education					
(Ref: Below Primary Level)					
Primary	0.966	0.146	6.63	0	0.680 – 1.252
Middle	1.781	0.126	14.14	0	1.535 – 2.028
Secondary	2.511	0.126	19.98	0	2.265 – 2.757
Higher Secondary	3.743	0.122	30.63	0	3.503 – 3.982
Graduate	4.360	0.122	35.88	0	4.121 – 4.598
Age	-0.008	0.001	-5.91	0	-0.011 – -0.005
Gender					
(Ref: Female)					
Male	-0.726	0.029	-25.14	0	-0.783 – -0.669
Religion					
(Ref: Hindu)					
Others	0.039	0.032	1.24	0.216	-0.023 – 0.102
Sector					
(Ref: Rural)					
Urban	0.304	0.028	10.79	0	0.248 – 0.359
Social Group					
(Ref: Others)					
ST	0.456	0.041	11.09	0	0.375 – 0.536
SC	-0.031	0.043	-0.71	0.476	-0.115 – 0.054
OBC	-0.196	0.032	-6.11	0	-0.259 – -0.133
Constant	-4.406	0.136	-32.45	0	-4.672 – -4.140

Table A3: Balance of Covariates- Formal versus Informal VET

Sample	PsR2	LRchi2	p>chi2	MeanBias	MedBias	B	R	%Var
Unmatched	0.231	11591.45	0	32.9	28.5	147.8*	0.56	100
Matched	0	3.63	0.989	0.6	0.4	3	1.09	0

Table A4: Logit model (Choice of Formal versus No VET)

	Coef.	Std. Err	z	P>z	[95% Conf. Interval]
General Education:					
Below Primary Level (Ref.)					
Primary	1.176	0.146	8.07	0	0.891 – 1.462
Middle	2.109	0.126	16.75	0	1.862 – 2.356
Secondary	2.759	0.126	21.97	0	2.513 – 3.005
Higher Secondary	3.986	0.122	32.65	0	3.747 – 4.225
Graduate	4.398	0.121	36.28	0	4.161 – 4.636
Age	-0.011	0.001	-8.5	0	-0.014 – -0.009
Gender:					
Female (Ref.)					
Male	-0.493	0.028	-17.63	0	-0.548 – -0.439
Religion:					
Hindu (Ref.)					
Others	0.126	0.031	4	0	0.064 – 0.188
Sector:					
Rural (Ref.)					
Urban	0.196	0.028	7.07	0	0.142 – 0.25
Social Group:					
Others (Ref.)					
ST	0.54	0.041	13.28	0	0.46 – 0.619
SC	-0.03	0.042	-0.72	0.473	-0.113 – 0.052
OBC	-0.048	0.032	-1.53	0.125	-0.11 – 0.013
cons	-4.694	0.135	-34.85	0	-4.958 – -4.43

Table A5: Balance of Covariates- Formal versus No VET

Sample	PsR2	LRchi2	p>chi2	MeanBias	MedBias	B	R	%Var
Unmatched	0.208	10401.28	0	30.2	26.9	138.8*	0.42*	100
Matched	0	2.35	0.999	0.5	0.3	2.4	1.04	0

Table A6: Logit model (Choice of Informal versus No VET)

Variable	Coef.	Std. Err	z	P>z	[95% Conf. Interval]
General Education:					
Below Primary level					
Primary	0.223	0.02	10.91	0	0.183 – 0.263
Middle	0.303	0.018	16.75	0	0.268 – 0.339
Secondary	0.218	0.021	10.51	0	0.177 – 0.259
Higher Secondary	0.192	0.022	8.85	0	0.15 – 0.235
Graduate	-0.005	0.021	-0.24	0.807	-0.047 – 0.037
Age	-0.002	0.001	-3.37	0.001	-0.003 – -0.001
Gender: Female					
Male	0.376	0.013	28.37	0	0.35 – 0.402
Religion: Hindu					
Others	0.154	0.014	11.25	0	0.127 – 0.181
Sector: Rural					
Urban	-0.125	0.012	-10.18	0	-0.149 – -0.101
Social Group: Others					
ST	0.032	0.019	1.66	0.096	-0.006 – 0.069
SC	-0.058	0.018	-3.22	0.001	-0.093 – -0.023
OBC	0.093	0.015	6.33	0	0.064 – 0.122
Constant	-0.373	0.033	-11.36	0	-0.437 – -0.309

Table A7: Balance of Covariates- Informal versus No VET

Sample	PsR2	LRchi2	p>chi2	MeanBias	MedBias	B	R	%Var
Unmatched	0.011	1890.88	0	6	4.8	24.7	0.89	100
Matched	0	6.5	0.889	0.3	0.3	1.4	1.02	0

Figure A1: Propensity score distributions of treated and control groups (Formal versus Informal VET)- Outcome variable: Daily earnings of regular or salaried employees

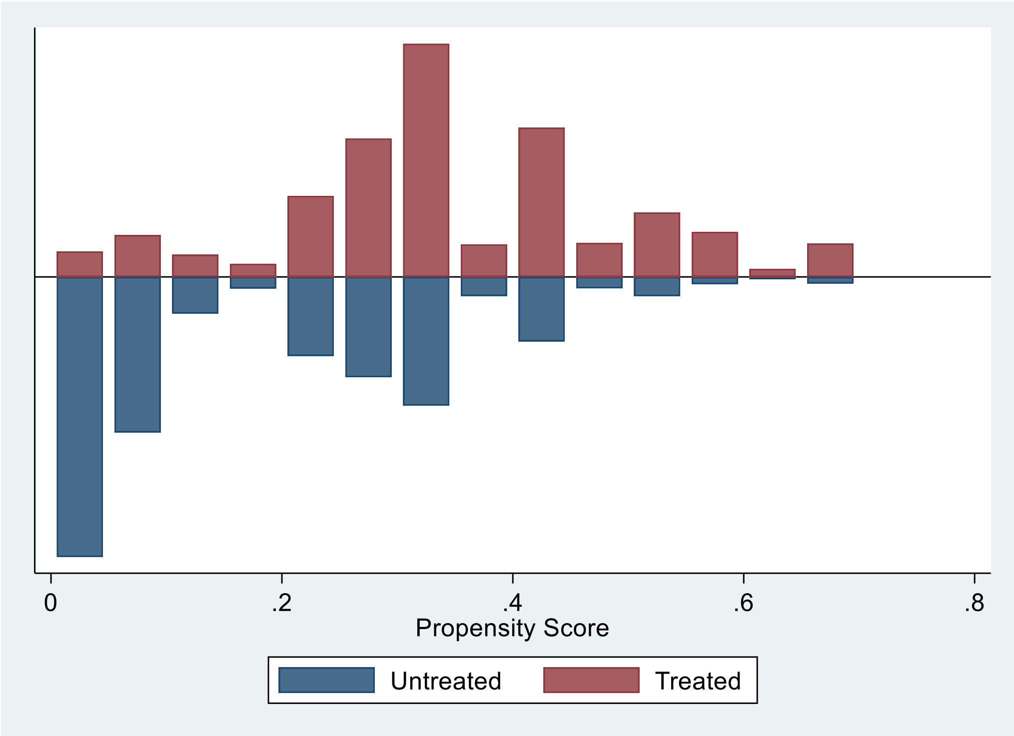


Figure A2: Propensity score distributions of treated and control groups (Formal versus No VET)- Outcome variable: Daily earnings of regular or salaried employees

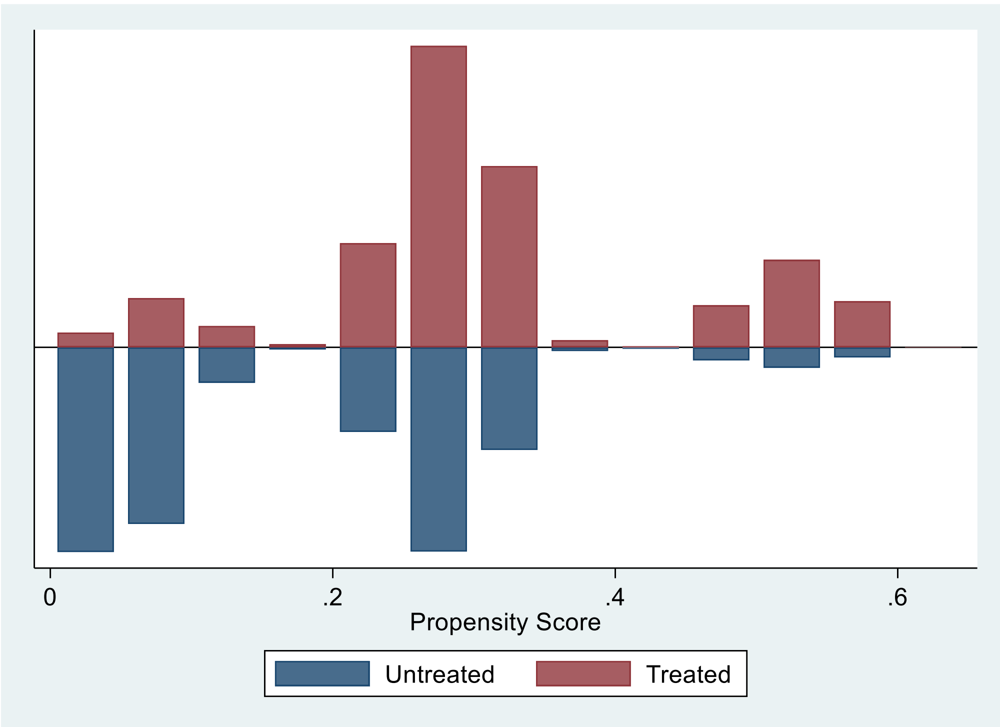


Figure A3: Propensity score distributions of treated and control groups (Informal versus No VET)- Outcome variable: Daily earnings of regular or salaried employees

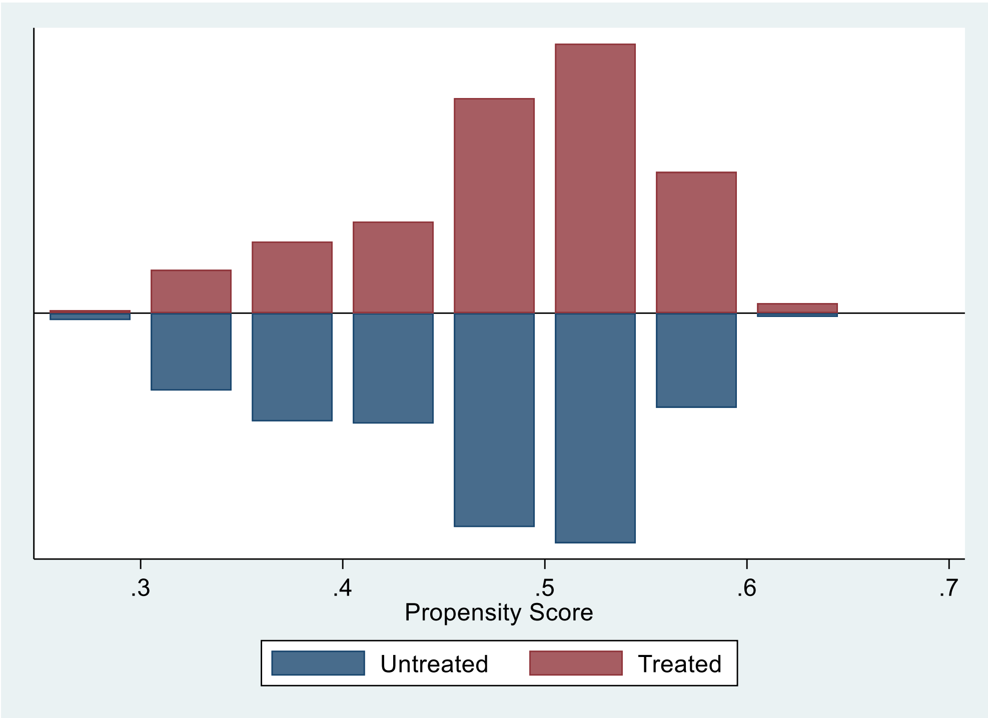


Figure A4: Propensity score distributions of treated and control groups (Formal versus Informal VET)- Outcome variable: Daily earnings of the self-employed

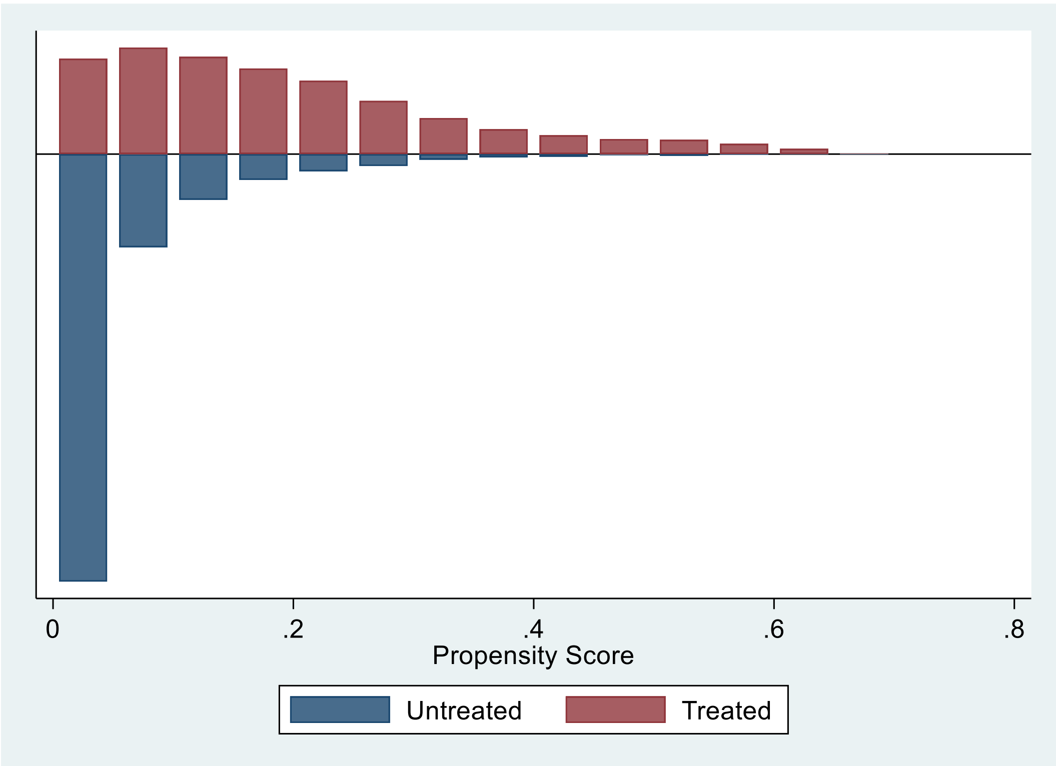


Figure A5: Propensity score distributions of treated and control groups (Formal versus No VET)- Outcome variable: Daily earnings of the self-employed

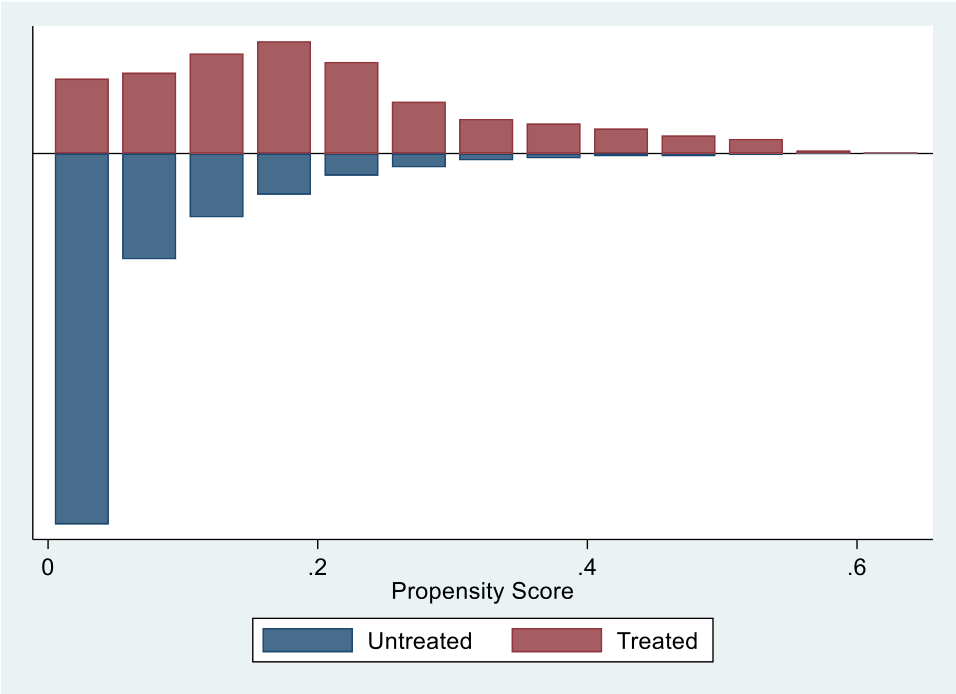


Figure A6: Propensity score distributions of treated and control groups (Informal versus No VET)- Outcome variable: Daily earnings of the self-employed

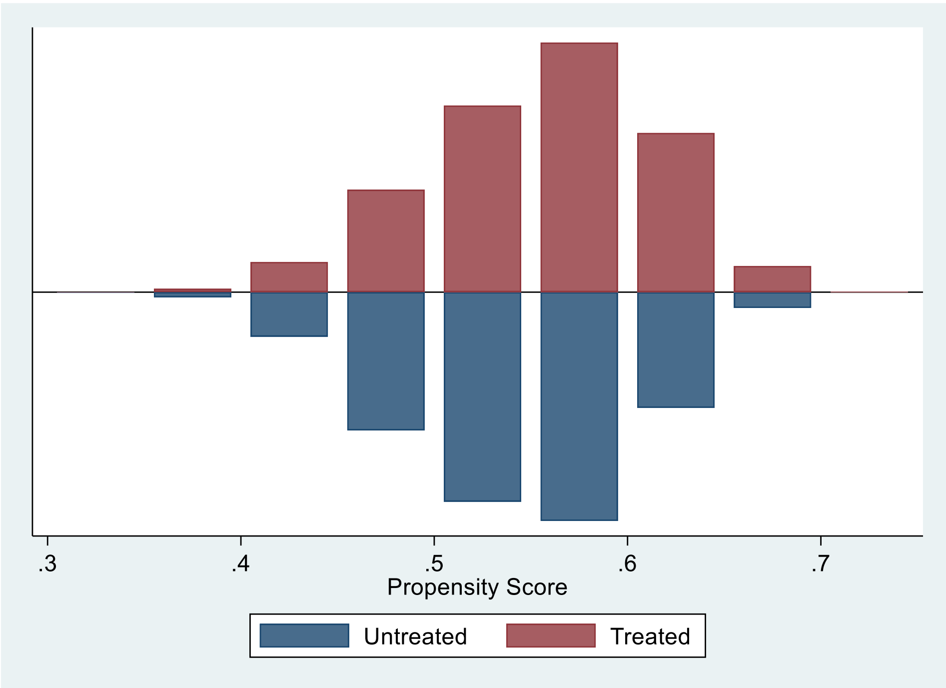


Figure A7: Comparison of propensity score before and after matching- Formal versus Informal VET (Earnings of regular or salaried employees)

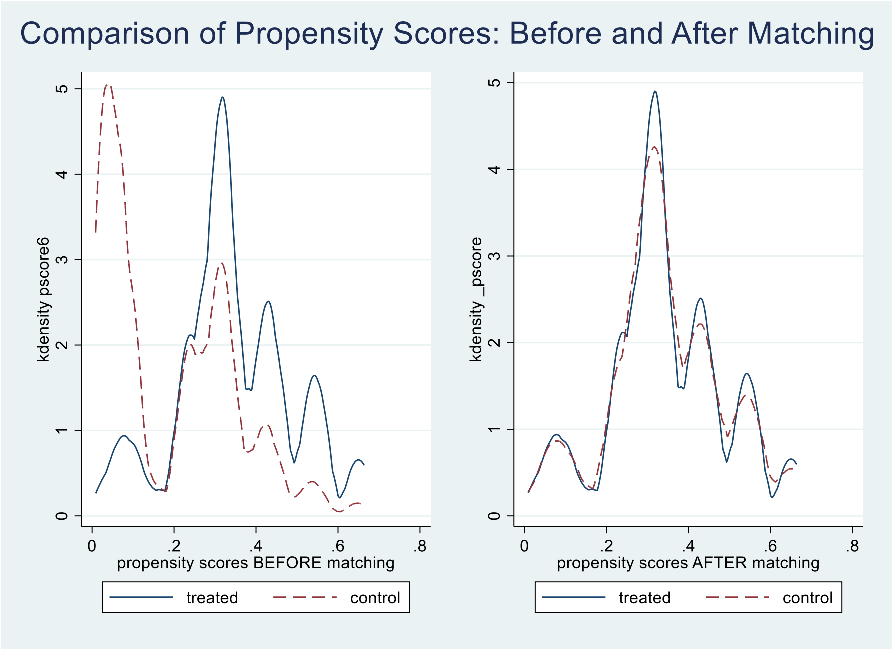


Figure A8: Comparison of propensity score before and after matching- Formal versus No VET (Earnings of regular or salaried employees)

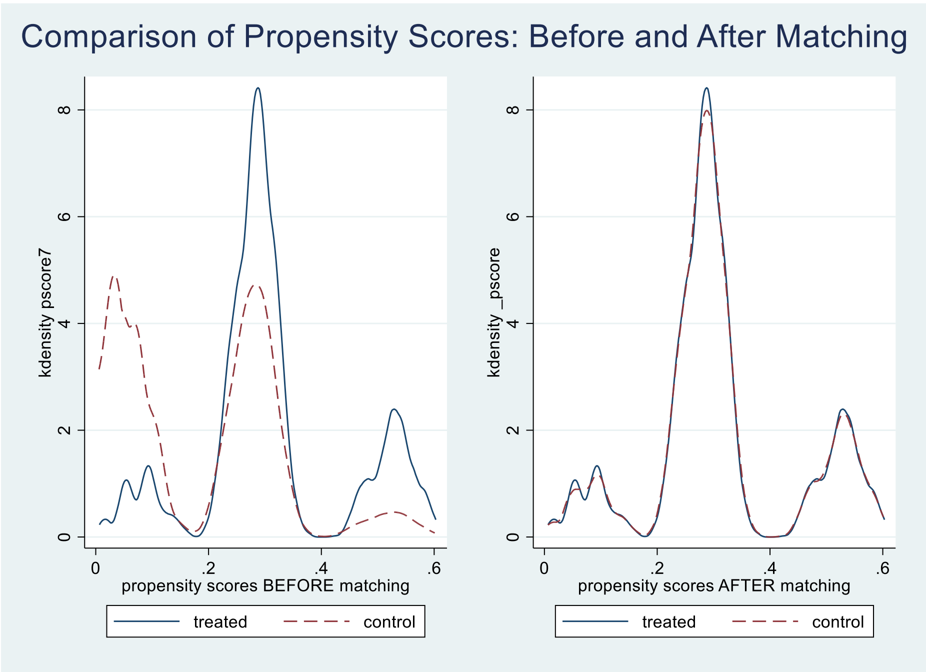


Figure A9: Comparison of propensity score before and after matching- Informal versus No VET (Earnings of regular or salaried employees)

Figure A10: Comparison of propensity score before and after matching- Formal versus Informal VET (Earnings of self-employed)

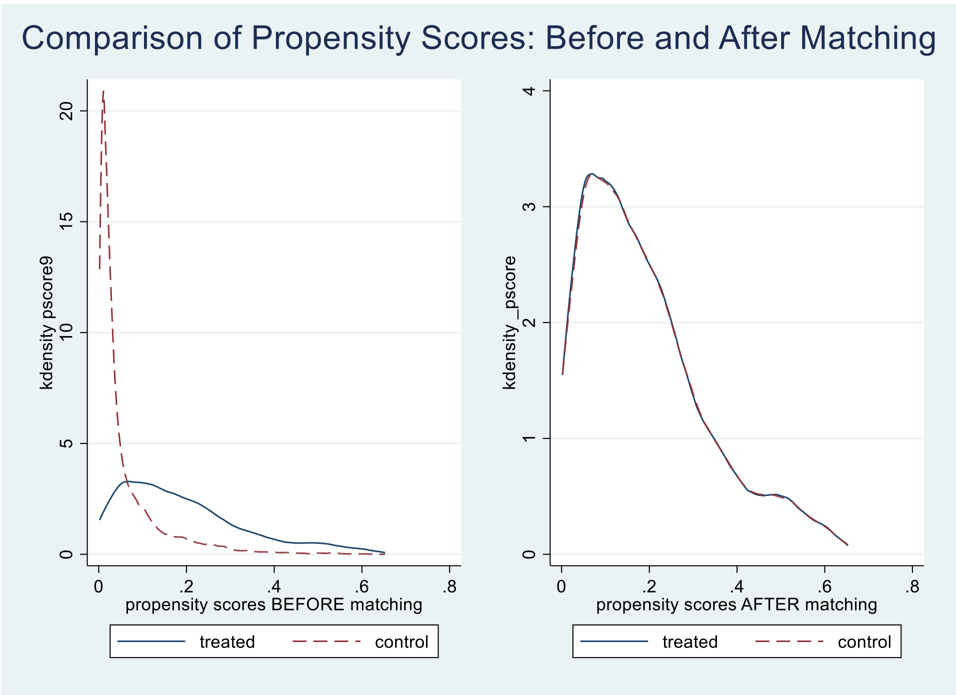


Figure A11: Comparison of propensity score before and after matching- Formal versus No VET (Earnings of self-employed)

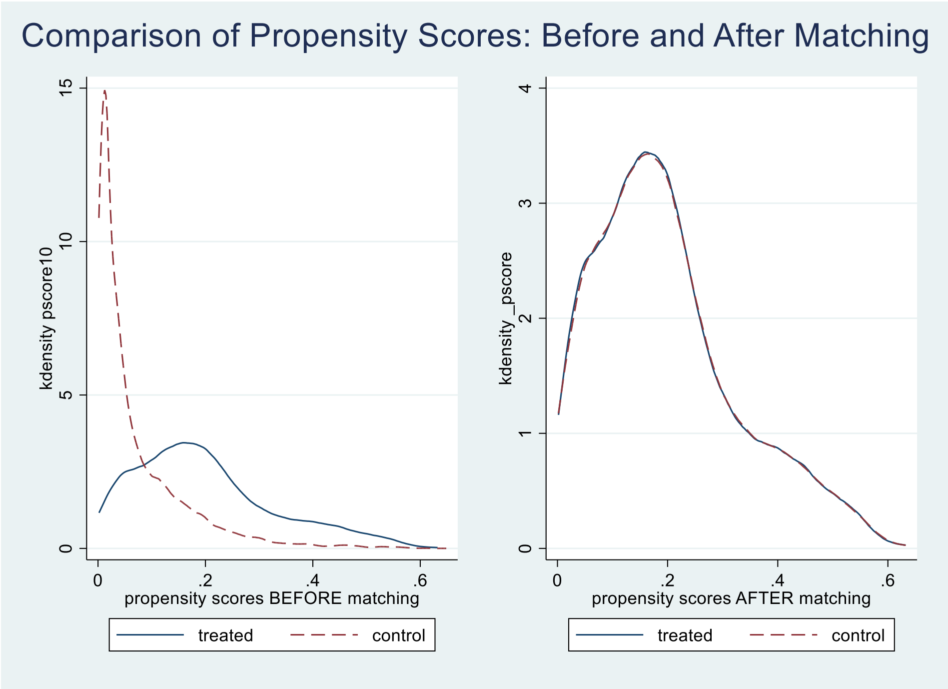


Figure A12: Comparison of propensity score before and after matching- Informal versus No VET (Earnings of self-employed)

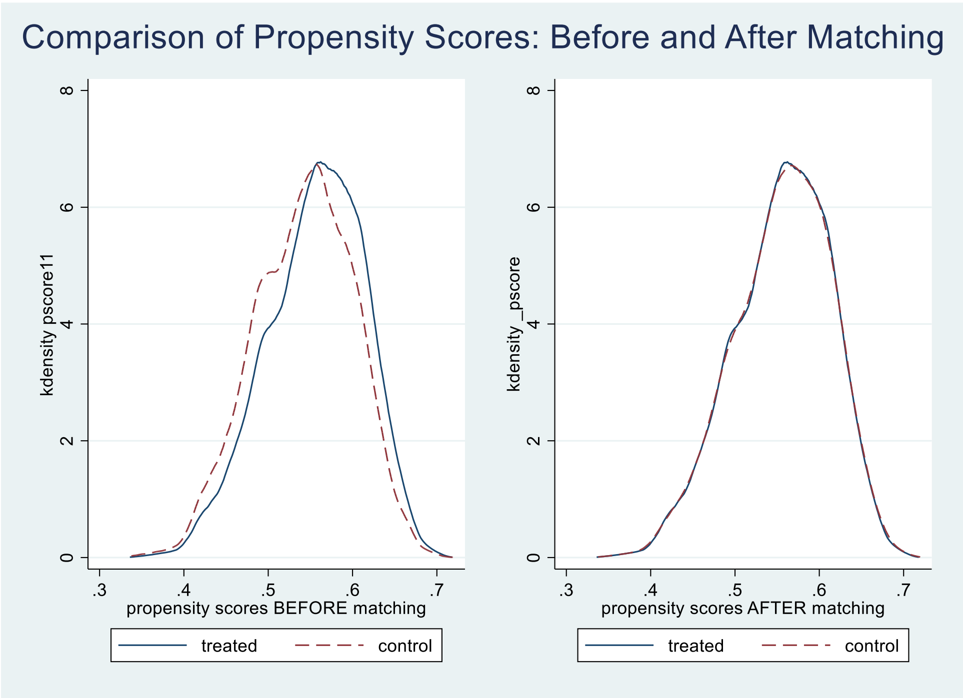


Table A8: Balance of Covariates- Propensity Score Matching

Model	Sample	PsR2	LRchi2	p>chi2	MeanBias	MedBias	B	R	%Var
Formal vs Informal VET: Earnings of regular/salaried employees	Unmatched	0.164	4240.96	0	27.8	22.2	110.1*	0.41*	100
	Matched	0	0.16	1	0.2	0.1	0.8	1	0
Formal vs No VET: Earnings of regular/salaried employees	Unmatched	0.144	3875.09	0	24.8	22.4	102.3*	0.36*	100
	Matched	0.001	9.4	0.669	1.4	1.1	5.9	1.1	0
Informal vs No VET: Earnings of regular/salaried employees	Unmatched	0.016	861.35	0	7.3	6.7	29.7*	0.84	0
	Matched	0.001	36.85	0	1.4	0.8	6.3	1.09	0
Formal vs Informal VET: Earnings of self-employed	Unmatched	0.202	3534.82	0	31.2	34.2	138.4*	0.84	100
	Matched	0.001	4.07	0.982	1.4	0.8	5.8	1.11	0
Formal vs No VET: Earnings of self-employed	Unmatched	0.187	3094.81	0	28.3	23.2	130.6*	0.64	100
	Matched	0.001	9	0.703	2	1.7	8.7	1.12	0
Informal vs No VET: Earnings of self-employed	Unmatched	0.01	820.42	0	5.6	5.1	23.8	0.98	0
	Matched	0	31.58	0.002	1	0.7	4.4	1.07	100