Supplying Vocational Training Institutes: Evidence from India

Aruj Shukla [*]

August 26, 2024

Abstract

Government across developing countries have addressed the jobs problem by investing in skilling and vocational training. Prior research has identified multiple demand-related channels to explain tepid returns to such investments. However, little is known about the supply of vocational training, especially the kind of trades offered and the quality of training provided. By constructing a novel dataset on the universe of vocational training institutes (VTIs) in India, I am able to establish five crucial stylized facts. First, I find that government training institutes perform 0.6 SD better than private counterparts based on a holistic quality assessment score. Second, better quality institutes tend be older, located in urban settings, run by the government, and offer more trade courses. Third, I find no preliminary support for competition driving up the quality of training provided. Fourth, in post economic liberalization India, the supply of private vocational institutes is responsive to literacy and population growth—in particular of Scheduled Castes. Finally, the types and array of trade courses offered remain unresponsive to evolving local economic and demographic conditions. Taken together, these results point to lack of dynamism in trades offered and private provision of VTIs as significant barriers to the supply of quality vocational training in India.

[*]University of Southern California. Email: arujshuk@usc.edu. I thank Vittorio Bassi, Julian Duggan, and Charity Troyer Moore for their guidance and productive discussions throughout the project.

1 Introduction

Skill acquisition is a key ingredient in not just realizing the potential of young job-seekers but also re-skilling experienced workers in any labor market. Technical and Vocational Education and Training (TVET) remains a critical channel for young, inexperienced workers to train for and meet demands of the evolving labor market. Labor markets in high income countries are testament to this avenue: about 52% of workers aged 19-24 in the USA; 72% in Germany; and 96% in South Korea have had some formal vocational training (UIS, 2016). Developing countries, however, have struggled to provide relevant and quality TVET in a fast-evolving landscape of skill requirements, demographic change, and deepening inequalities. A steady growth in youth unemployment rates is a reflection of this deficiency that calls upon policy makers to act expeditiously in addressing the issue. [1]

Tackling the jobs problem in poorer countries requires both demand- and supply-side solutions. Jensen (2012) highlights the vital role for creating employment opportunities that incentivize investments in skills. Supplementing this demand-side push, the provision of high quality skills that are accessible to vast segments of the workforce is equally crucial Much like in health and education sectors, developing countries have made tremendous strides in expanding the availability of vocational training institutes.[2] A combination of private markets and the public sector have led the expansion, adding to the perennial discussions on which sector should provide these services (Coarasa et al., 2014). Regardless of the source—public or private—of provision, ultimately, it is the quality of services that bears paramount importance (Das and Hammer, 2014; Andrabi et al., 2015).

In this paper, I study the supply of vocational training institutes (VTIs) in India. First, I summarize some basic facts about vocational training institutions with an emphasis on variation in spatial location, access to a VTI, and the set of trade courses offered. Next, I bring evidence to the quality of vocational training across privately- and publicly-run institutes, highlighting differences and general correlates of quality in the Indian context. Finally, I assess the variation in the entry of VTIs and, conditional on entry, the type of trades offered over the past three decades. This analysis underscores the importance of dynamism in TVET curriculum setting and the role of local demographic and economic conditions in shaping the supply of quality vocational training.

Through a series of web scrapes, I assemble a novel dataset on the universe of 14,951 VTIs in India. Mapping these, I find government VTIs are clustered in traditionally industrial states and yet have good coverage in remote districts across the country. Private VTIs concentrate around major cities and, expectedly, have lower coverage than government ones, with notable absence in the north eastern states. Using scores from a recent government-led grading initiative that assigned grades to all VTIs in the country, I find that government run VTIs outperform their private-run counterpart in terms of quality of training. Controlling for a variety of characteristics, I find private VTIs to be 0.6 SD units worse in quality than government ones. Higher quality institutes offer more trades, are older, and are located in urban areas. Interestingly, I find little evidence that competition—measured by density of VTIs in a district—is associated with higher quality. Finally, I combine data from three waves of the Population Census and four waves of the Economic Census to explore variation in the entry of VTIs and trades offered within a district over time. Increase in population—especially that of marginalized subgroups—and literacy correlate with more VTIs, particularly private institutes. Employment statistics exhibit a muted association.

[1] According to International Labour Organization (2022), globally, youth unemployment was 16% in 2022.

[2] Brazil, India, Kenya, and Philippines are some notable examples.

Most notably, the range of trades and types of trades (engineering and non-engineering trades) are not sensitive to local demographic and economic conditions. The ossification in curricula offered reinforces prior qualitative work and warrants further scrutiny, which I discuss briefly for future work.

This paper contributes to building an understanding of the supply side determinants of vocational institutions, which has not been explored thoroughly in the literature. From the entry decision and location choice to the personnel retention policies and operational costs, we know little about the quality gradient of VTIs and how profitable they are. As a starting point, I present a series of descriptive statistics and correlations to shed light on the vocational skill acquisition landscape in India.

In doing so, I attempt to make two distinct contributions. First, I supplement the existing research on low takeup of vocational training by providing a supply-side perspective. Current work finds mismatched wage expectations and social norms as key bottlenecks on the demand side. Prillaman et al. (2017) find that Indian women face familial pressure and mobility constraints that prevent them from gaining vocational training. Discouragement stemming from high aspirations could be another drag on enrollment (Genicot and Ray, 2020; Kumar, 2022). On the supply side, the lack of industry linkages, quality of training and management, and legacy curriculum are cited as constraints. This paper identifies a discernable quality gradient, which in conjunction with asymmetric information between provider and candidate has the potential to be a significant driver in the lackluster performance and take up of VTIs (Rageth and Sritharan, 2022). In addition to the focus on quality differences across public and private provision of VTIs, I also add empirical, policy-relevant evidence to questions of financing, provision, and regulation of essential goods and services. The push to either expand the public provision of vocational training or institute tighter regulation on the provision of private institutions will ultimately depend on the quality differential across sectors—a domain this paper can add value to (Bank, 2022; Coarasa et al., 2014; NITI Aayog, 2023).

2 Background

2.1 Vocational Training in India

Since the 1980s, the Indian government has undertaken several initiatives to expand and enhance vocational training programs. The National Policy on Education (1986) marked an early push by integrating vocational education into education policy. The 1990s witnessed an expansion of VTIs and the establishment of the National Council for Vocational Training (NCVT) to set standards and regulate aspects of the education and training provided. Since the turn of the millennium, the creation of the National Skill Development Corporation (NSDC) marked the shift to fostering skill development through public-private partnerships. As of FY 2023-24, $464 million was allocated to TVET in the national budget.

Despite continued attempts at financing, provision, and regulation, only 2.4% of Indian workers have technical education of any kind (Mehrotra et al., 2020). Issues such as a mismatch between the skills provided and industry requirements, disparities in training access, and the need for updated curricula to keep pace with technological and skill advancements persist. Moreover, improvements in infrastructure and the supply of qualified instructors is a key bottleneck, resolving which can bridge the oft cited gap between theoretical skills and practical needs of the candidates (Mehrotra et al., 2020; NITI Aayog, 2023).

3 Data & Outcomes

3.1 Sources

To analyze the supply side of the vocational training landscape in India, I use a combination of publicly available administrative data and novel data on VTIs.

Administrative Data The administrative data for the analysis comes from waves of the decennial Population Census (1991, 2001, 2021) and the Economic Census (1990, 1998, 2005, 2013). I obtain a host of basic demographic and economic variables ranging from population, number of literate individuals, and total employment, including disaggregated numbers by broad sectors (manufacturing, services, informal) of the economy. The key challenge with using multiple rounds and types of census data is the stability of the geographical unit. Asher et al. (2021) present a thorough and open-source solution to this issue by making available The Socioeconomic High-Resolution Rural-Urban Geographic Dataset on India (SHRUG) to researchers. SHRUG is an open dataset that collates administrative data from multiple government sources and links them over time and across sources. Through a granularly defined ID (`shrid`)—at the town or village level—I first assemble a panel at this level of granularity and then aggregate up to the district level.

Shapefiles Although SHRUG provides shapefiles of the administrative polygons in India, the most recent one is dated to 2016. As districts boundaries across India can change frequently, for the mapping exercise, I use for the most recent shapefile data from 2022. [3] There are 741 districts in this dataset and the maps represent the as-is spatial landscape of the vocational training sector—concordant with the data on VTIs described below. I use older boundaries for the regression analysis as merging is less tricky and allows for consistent datasets across the cross sectional and panel analysis.

VTI Data To the best of my knowledge, there is no publicly available dataset on VTIs in India. Putting together disparate data sources, I construct a novel, cross-sectional dataset on the universe of VTIs by scraping the National Council for Vocational Training's (NCVT) Management Information System (MIS). [4] Key information available includes basic details of the institution (name, location, date of establishment, type of institution (public/private) etc) along with a list of trades offered and affiliation details of the VTI. I also scrape the grade details of a recent initiative undertaken by the Ministry of Skill Development & Entrepreneurship (MSDE) to assign a quality score to every VTI in the country based on key performance parameters. [5] The grades are available for $\approx$ 98% of the 14,951 VTIs in India. Finally, using fuzzy matching algorithms on a list of trades from the Directorate General of Training (DGT) website, I classified the trades offered at every VTI into two broad categories: Engineering and Non-engineering trades. Welder, tool and die maker, and electrician are some popular examples of engineering trades; computer operator, sewing, and dress-making are examples of non-engineering trades. The resulting dataset is rich with key variables that I describe next.

[3] I source the shapefiles from this GitHub repository.

[4] See here for an example of a VTI in the state of Andhra Pradesh. I scrape all the information on this page for all Indian VTIs.

[5] See here for further details on this initiative.

3.2 Constructing a district × year panel

In order to understand the temporal variation in the prevalence of VTIs and in the array of trades they offer, I combine VTI data and Census data to construct a panel at the district × year. My choice of district as the level of aggregation is rooted in the observation that a district is populous enough to capture meaningful variation in the number of VTIs over time and across space, without posing a challenge with inference.[6] The average Indian district in 2011, the most recent population census year, was home to roughly 2 million people and there were 640 districts in India.

To convert the cross-sectional data on VTIs into a panel, I rely on the location and year of establishment of the VTI to denote whether an institution was operational in a certain year. Aggregating to the district level yields the stock of VTIs at a district × year level. In similar fashion, I calculate the stock of unique trades offered at a district in a given year. As I don't observe the starting year for a particular trade offered at any VTI, it is important to note that any temporal change in the number of unique trades offered in a district can only be due to the entry of a new institution that offers a completely different trade.

Drawing from the waves of the Population Census and the Economic Census warrants an additional step of aligning years across sources when constructing the panel. Assuming a constant growth rate of population between the decennial population census waves, I extrapolate values from the 1991, 2001, 2011 Population Census to the Economic Census years (1990, 1998, 2005, 2013). For instance, I extrapolate the 2001-2011 growth rate to get the population for 2005 and 2013—years for which there are no actual demographic statistics available. The resulting panel spans 7 census years and has data on 591 of the 640 Census 2011 districts. [7]

3.3 Outcomes

The primary outcomes of interest are the number of VTIs and the number of unique trades offered. I also construct an extensive margin analog of these variables: whether a new VTI opened in a district and whether an engineering or non-engineering trade was offered in the district. Another important variable—both as an outcome and explanatory variable—is the overall grade received by the VTI. Consisting of eight parameters the VTI score ranges from 0 to 10 points. [8] I further standardize this score with a zero mean and standard deviation of one. [9] In analyses comparing private and public VTIs, I construct the standardized scores with respect to the distribution of public VTI scores such that any change can be interpreted relative to scores of public VTIs.

Finally, as the set of regressors, I use the population and economic census variables defined above. To aid interpretability of the coefficients, I rescale the census regressors by 10,000. A unit change then signifies an economically and demographically significant change, making the results easier to assess.

[6] At a sub-district or village level, I am unlikely to observe multiple vocational institutes, which precludes the addition of fixed effects. I plan to add this level of analysis in future workstreams.

[7] Based on the Census 2011 district polygons and the VTI geo-coordinates available, the remaining 49 districts don't appear to have any vocational institution and are excluded from the analsysis.

[8] The parameters include Minority group (SC/ST/PwD) enrollment (5%), Female participation (5%), Trade Diversity (5%), Dual System of Trainee enrollment (5%), Overall admission percentage (35%), Pass rate (35%), Past trainee grades (5%), and Computer-based test participation (5%)

[9] I construct the standardized score by netting out the sample mean and then dividing by the sample standard deviation.

4 Empirical Strategy

I proceed with the analysis in three distinct steps, each of which requires a different empirical strategy and builds on the prior set of analysis. The spirit of the exercises is to isolate the determinants of the evolution of the number of VTIs and the trade courses they offer over the past few decades. The distinction between public and private run VTIs is not just interesting from an analytical standpoint but economically meaningful too. Wherever possible, I implement the following analysis on the sub sample of government and private VTIs separately.

4.1 Mapping exercise

To start with, I conduct a basic mapping exercise that visualizes the location of every VTI in the country as of 2024. [10] As a descriptive exercise, the goal is to shed light on the current state of VTIs in India. From highlighting the spatial variation in the location of VTIs to the range of trades accessible to individuals in their vicinity, this exercise sets the foundation for further analyses.

4.2 Cross sectional analysis

To add structure to the variation observed in the prior exercise, I then turn to the cross sectional data on the universe of VTIs. Leveraging a vector of VTI level details, I independently investigate the correlates of the quality of the VTI by implementing the following specification:

$$score_{is} = \beta_1 characteristics_{is} + \alpha_s + \varepsilon_{is} \tag{1}$$

where $score_{is}$ is the quality grade received by VTI i in state s and α_s denotes state fixed effects that I include to absorb significant variation across states in India. I cluster the standard errors at the district level. [11]

Building on this specification, I explore the correlates of the number and types of trades offered at a VTI. Specifically, I estimate:

$$y_{ids} = \beta_1 characteristics_{ids} + \gamma_d + \alpha_s + \varepsilon_{ids} \tag{2}$$

The key distinction from 1 is the addition of γ_d as a control, which is defined as the number of VTIs in a district per million residents and is a preliminary measure of competition. Accounting for the number of such institutions can paint a clearer, albeit not causal, determinant of the choice of trades to offer.

4.3 Panel analysis

In this final exercise, I use the district $\times$ year panel to understand the changes in the entry of VTIs and subsequently the trades they offer as a function of the local demographic and economic conditions. Using a set of district and time fixed effects, I estimate:

$$y_{dt} = \alpha_d + \beta_1 CensusVar_{dt} + \gamma_t + \varepsilon_{dt} \tag{3}$$

[10] Latitude/Longitude coordinates for 142 of the 14,951 VTIs are missing. At best, by inferring the other set of details I scraped, I can map them to the state they are located in. For now, I exclude these VTIs from the analytical dataset.

[11] In India, administrative divisions, typically, adhere to the following hierarchy: a state is made up of districts. Districts constitute sub-districts or blocks, which are made up of a set of villages.

where α_d and γ_t are district and time fixed effects respectively. Here, β_1 picks up the variation within the district over time. Standard errors are clustered at the district level. Additionally, I extend this specification by adding the vector of census variables as the regressor set. β_k then summarizes the partial association between regressor k and the outcome variable in a district over time, holding all else equal.

5 Results

5.1 Mapping exercise

I first present a map of every (mappable) VTI in India in Figure 1. The density of privately run VTIs (in blue) and the geographic spread of government run VTIs (in red) is staggering. Private VTIs continue to be the most available option for prospective students; however, government VTIs fill provision gaps in hard to access areas where the private enterprise is not profitable and thus absent.

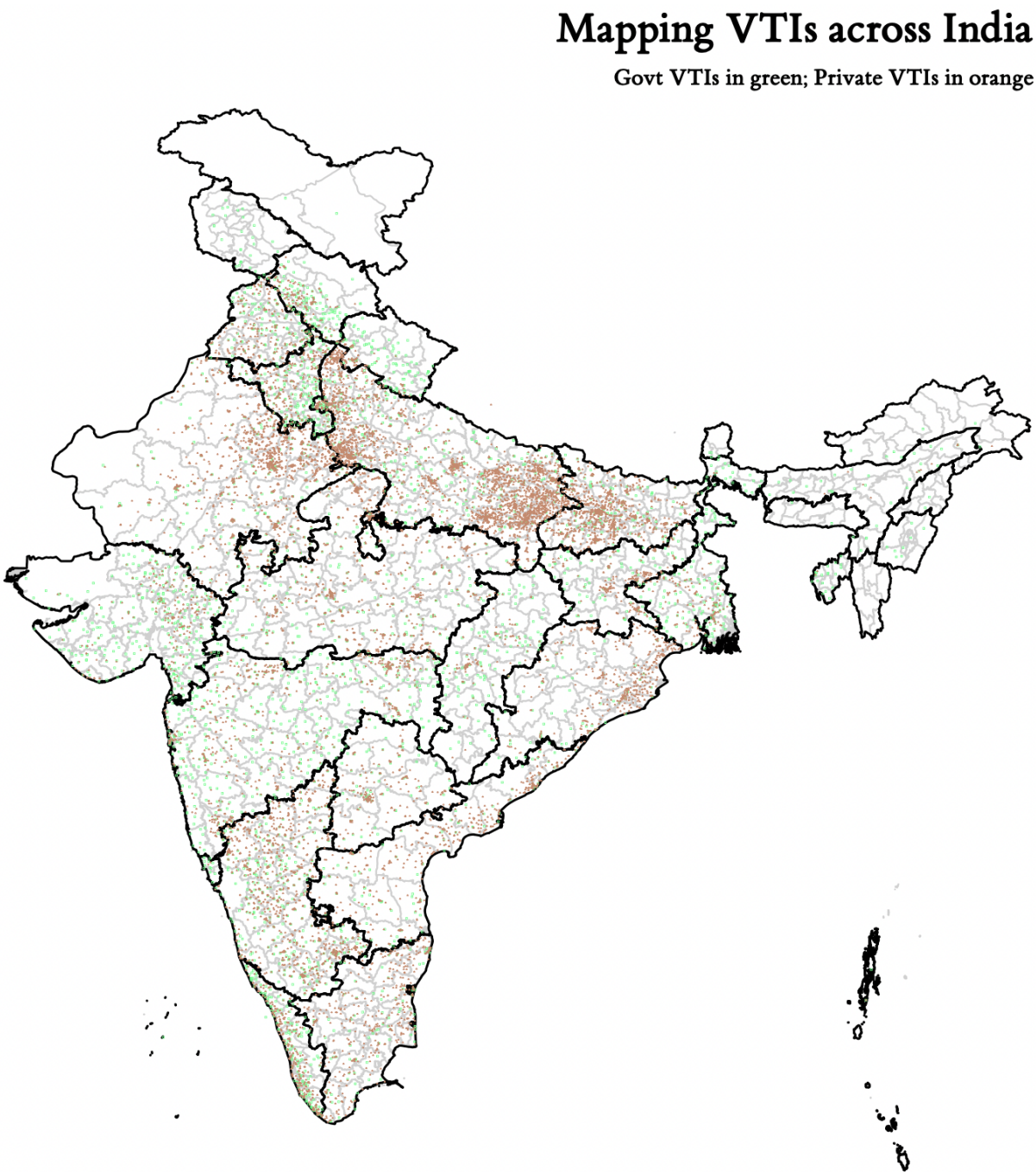


Figure 1

For more concrete insights, Figure 2 visualizes the current number of private and government VTIs in a district across India. This choropleth map provides some striking insights. Government VTIs are highly concentrated in the south western parts of India (states of Gujarat, Maharashtra, and Karnataka), which have traditionally been industrial hubs. Despite this regional concentration, government VTIs have better penetration compared to private ones, which are all but absent in the north eastern states and in Kashmir. Private VTIs appear to be highly clustered in two notable regions. The first is a large radius around the capital city state of Delhi and spans mostly rural districts of the north western states of Haryana and

Rajasthan. Proximity to dense urban location where most jobs are situated is a likely explanation for this cluster. The other cluster straddles the eastern border of Uttar Pradesh and Bihar—both states that are marked by high levels of poverty and low literacy rates. Despite these drawbacks, the sheer enormity of the population and the consequent volume of goods and services produced is likely a noteworthy factor in determining the demand for VTIs.

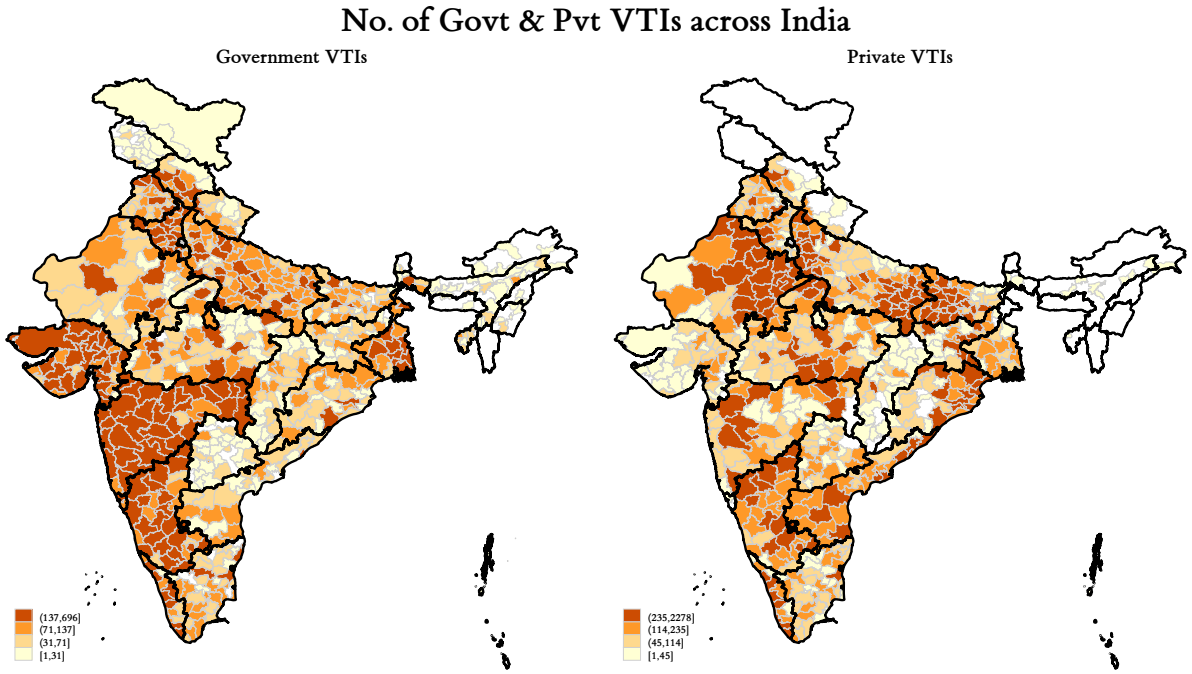


Figure 2

Beyond the presence of VTIs in a district, I also investigate the number of unique trades that are offered in a district. A new, in-demand trade offering can be a source of differentiation and a robust indicator for the quality and competitiveness of the institution. It also provides a snapshot of the specialization in training that's available in the students' immediate vicinity. Figure 3 presents the results of this exercise for private and government VTIs alike. Conditional on the presence in a district, private VTIs seem to be aligned with government VTIs in the number of trades they offer; the western state of Maharashtra provides the perfect case study. Overall, at this level of aggregation, there appears to be a great deal of persistence and competition in the number of trades offered. The latter is welcome but the former may indicate a lack of dynamism in the curriculum that can have a knock-on effect on the supply of a skilled labor force.

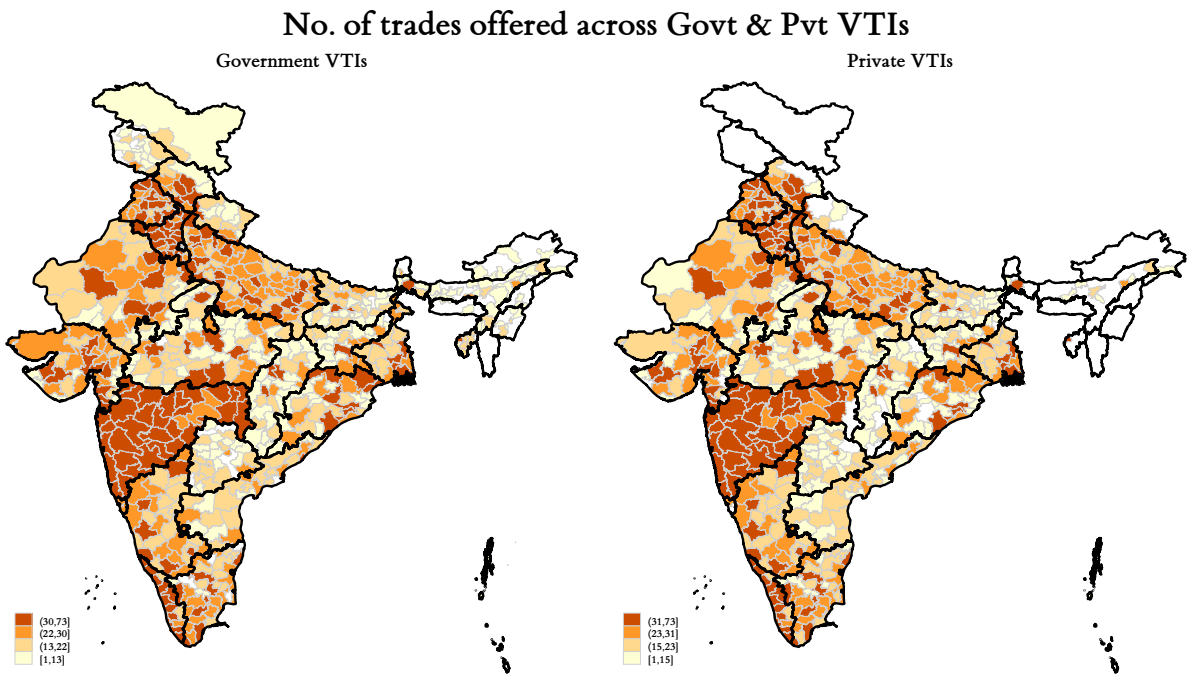


Figure 3

5.2 Cross sectional regressions

The mapping exercise presented some interesting insights that I explore further in this section. Of particular interest is the fact that private VTIs cluster heavily in populous regions and offer little variation in the breadth of trades offered. Such a setting brings the quality of the VTI to the forefront: a dimension for differentiation in a competitive market.

To understand the drivers of quality, I first construct a standardized version of the raw overall scores assigned to every VTI. These scores come from a recent (FY 2023/24) grading cycle and are a reasonable summary statistic for the quality of the institution. Table 1 presents a set of bivariate regressions of the standardized quality score on a series VTI level characteristics. I find no evidence for the hypothesis that competition enhances quality. If anything, using a basic measure of competition that takes the number of VTIs per (million) capita in a district, I find a slightly negative and noisy correlation with quality. Older VTIs, in urban locations, and offering more trades are associated with higher quality. Importantly, privately run VTIs perform, on average, 0.6 SD worse than the government ones. Figure 4 provides further visual validation of this finding: government VTIs are unambiguously superior in quality compared to private ones. Using the Kolmogorov-Smirnov test, I reject the null of equality of score distributions across government and private VTIs. It is noteworthy that while both types of VTIs have low-performing institutes, the mass of "superstar" institutes is dominated by government run VTIs.

Table 1: Correlates of VTI quality

	Standardized VTI quality score				
	(1)	(2)	(3)	(4)	(5)
No. of VTIs per million people	-0.012				
	(0.021)				
Years since establishment		0.016***			
		(0.001)			
Private VTI			-0.597***		
			(0.042)		
In urban location				0.066***	
				(0.025)	
No. of trades offered					0.042***
					(0.002)
Dep. variable mean	0.01	0.01	0.01	0.01	0.01
Observations	14478	14013	14559	14559	14559
No. of districts	569	575	578	578	578
No. of years	-	-	-	-	-
R^2	0.12	0.16	0.17	0.12	0.15

NOTES—Table above presents results from bivariate regressions in a cross-sectional dataset. Dependent variable is a standardized version of the overall VTI score, representing quality. Specification includes state fixed effects. Standard Deviation reported in square brackets and standard errors in parentheses. Standard errors clustered at the district level. *p<0.1, **p<0.05, ***p<0.01.

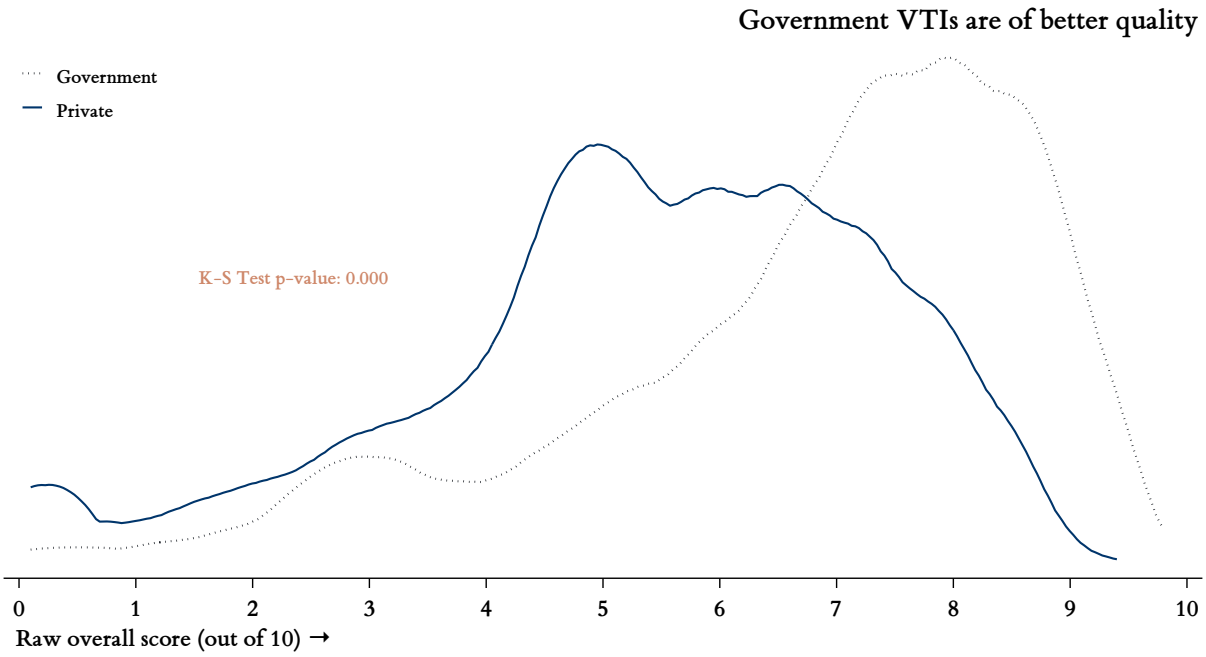


Figure 4

The outperformance of private VTIs runs counter to theories of market based provision of goods and

services. Efficient provision of in-demand services is the hallmark of well functioning markets. I aim to add nuance to this finding by taking a modular approach: a government run VTI is no more than a bundle of attributes including, but not limited to, factors such as years in operation, location, management, and funding. In the analysis that follows, I decompose the differences across private and government VTIs.

Figure 5 charts the recent explosion in the establishment of private VTIs. Since 2006, privately run VTIs have significantly outpaced the entry of government VTIs, resulting in a 80% share of all such institutes in the country.

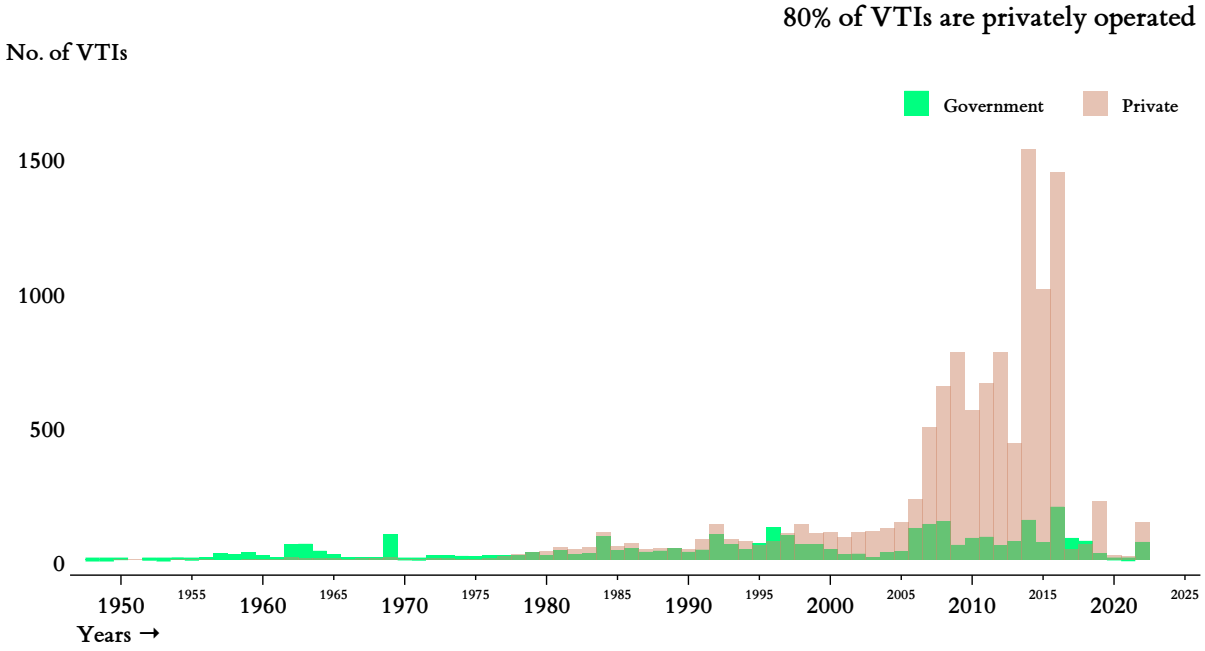


Figure 5

Table 2 summarizes other dimensions distinguishing private from government VTIs. Not only are private VTIs newer, they are more densely located in a district. Moreover, private VTIs are 3.5 percentage points less likely to be in an urban location and 42.5 percentage points less likely to offer a non engineering trade.[12] Taken together, private VTIs seem to cater to rural demand—likely due to cheaper land prices and favorable demographics—with lower quality institutions and offer a high proportion of engineering trades. A critical factor is the nature of demand for VTIs in rural areas. Learning engineering oriented trades can be a good skill to acquire but if there are significant migratory frictions or the demand for such trades is weak in the local economy, returns to this investment are likely to be negligible.

[12] Including district fixed effects allows for better comparison and I am able to reproduce these patterns. See Table A6 for details.

Table 2: Cross-sectional variation across private- and government-run VTIs

	Government ITIs (1)	Difference (2)	N (3)
No. of VTIs (in district)	33.326 [32.856]	16.119*** (1.893)	14806
No. of trades offered	8.494 [6.394]	-5.799*** (0.147)	14806
Years since establishment	28.033 [18.874]	-10.938*** (0.589)	14260
In urban location	0.283 [0.450]	-0.035*** (0.013)	14806
Engineering courses offered	0.954 [0.210]	-0.014** (0.006)	14806
Non-engineering courses offered	0.746 [0.435]	-0.425*** (0.017)	14806
Standardized VTI quality score	0.000 [1.000]	-0.632*** (0.044)	14559

NOTES—Specification includes state fixed effects. Standard Deviation reported in square brackets and standard errors in parentheses. Standard errors clustered at the district level. *p<0.1, **p<0.05, ***p<0.01.

Given the importance of the choice and range of trade offerings at a VTI, I regress this set of outcomes—independently for private and government VTI subsamples—on the quality score, age of the institute, and its urbanicity, while controlling for the density of VTIs in the district. Tables 3 and 4 summarize the results. While there is no association between quality and the offering of non-engineering trades in private VTIs, only the good quality government ones offer such trades. Across both types, older and urban VTIs are positively associated with such offerings. Higher quality government VTIs tend to offer more trades but higher quality private VTIs buck the trend by offering fewer. A possible explanation could be trade specialization but the overall lower quality of private VTIs complicates the narrative. Finally, I find no difference across the range of trade offerings across rural and urban locations for private institutes. This is in stark contrast to government run VTIs, which offer more trades in urban locales—in line with a larger variety of sectors in urban areas.

Table 3: Private VTIs: Correlates of trades offered

	No. of trades			Engineering trades			Non-engineering trades		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Standardized VTI quality score	-0.061***			0.000			-0.010		
	(0.021)			(0.003)			(0.007)		
Years since establishment		0.052***			-0.001***			0.003***	
		(0.004)			(0.000)			(0.001)	
In urban location			0.015			-0.034***			0.068***
			(0.043)			(0.005)			(0.013)
Observations	11384	11073	11557	11384	11073	11557	11384	11073	11557
R^2	0.16	0.24	0.16	0.08	0.06	0.08	0.14	0.13	0.14

NOTES—Table summarizes results from a set of regressions using a VTI level dataset where each point estimate reported is a separate regression. Outcome in columns (4)-(6) is an indicator for any engineering trade offered at VTI and the outcome in columns (7)-(9) is an analog for non-engineering trade offerings. Specification includes a control for the number of VTIs per million capita in district (as a proxy for competition) and state fixed effects. Standard errors are clustered at the district level *p<0.1, **p<0.05, ***p<0.01.

Table 4: Govt VTIs: Correlates of trades offered

	No. of trades			Engineering trades			Non-engineering trades		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Standardized VTI quality score	0.964***			-0.008			0.124***		
	(0.149)			(0.007)			(0.011)		
Years since establishment		0.197***			0.000*			0.004***	
		(0.009)			(0.000)			(0.001)	
In urban location			2.461***			-0.018**			0.040**
			(0.264)			(0.008)			(0.016)
Observations	3097	3113	3171	3097	3113	3171	3097	3113	3171
R^2	0.13	0.41	0.15	0.17	0.19	0.17	0.39	0.37	0.35

NOTES—Table summarizes results from a set of regressions using a VTI level dataset where each point estimate reported is a separate regression. Outcome in columns (4)-(6) is an indicator for any engineering trade offered at VTI and the outcome in columns (7)-(9) is an analog for non-engineering trade offerings. Specification includes a control for the number of VTIs per million capita in district (as a proxy for competition) and state fixed effects. Standard errors are clustered at the district level *p<0.1, **p<0.05, ***p<0.01.

5.3 Panel regressions

The panel structure of Census waves presents an opportunity to examine the evolution of VTIs in a district over time. The changing local conditions—both economically and demographically—may play a critical role in not just the entry of VTIs but also the trades they offer. Tables 5 and 6 present correlations between economic conditions and the variation in number of VTIs in a district. While there is a strong correlation with different employment indicators across the private and government institutes, the magnitudes present some contrast. The number of private VTIs respond positively and much stronger to employment changes, especially services and manufacturing, than government VTIs. Moreover, private institutes negatively associate with informal employment: an increase of 10,000 informal employees is associated with one fewer VTI in the district. With the prevalence of informality in India, this is a significant magnitude.

Table 5: Economic determinants of: number of Govt VTIs

	No. of Govt. VTIs				
	(1)	(2)	(3)	(4)	(5)
Total employment	0.566***				
	(0.204)				
Informal employment		0.020			
		(0.104)			
Private employment			0.586***		
			(0.214)		
Manuf. employment				1.070*	
				(0.607)	
Services employment					0.845***
					(0.282)
Dep. variable mean	75.96	75.96	75.96	75.96	75.96
Observations	2271	2271	2271	2271	2271
No. of districts	582	582	582	582	582
No. of years	4	4	4	4	4
R^2	0.95	0.95	0.95	0.95	0.95

NOTES—Table presents results from a set of bivariate regressions. Each row displays results from a separate regression. Specification includes district and year fixed effects. Standard errors clustered at the district level. Employment statistics refer to count of people employed. Row variables are reported in 10,000s. *p<0.1, **p<0.05, ***p<0.01.

Table 6: Economic determinants of: number of Private VTIs

	No. of Private VTIs				
	(1)	(2)	(3)	(4)	(5)
Total employment	1.811***				
	(0.449)				
Informal employment		-1.197***			
		(0.420)			
Private employment			1.947***		
			(0.491)		
Manuf. employment				2.683**	
				(1.255)	
Services employment					2.844***
					(0.672)
Dep. variable mean	51.41	51.41	51.41	51.41	51.41
Observations	2271	2271	2271	2271	2271
No. of districts	582	582	582	582	582
No. of years	4	4	4	4	4
R^2	0.63	0.62	0.63	0.62	0.63

NOTES—Table presents results from a set of bivariate regressions. Each row displays results from a separate regression. Specification includes district and year fixed effects. Standard errors clustered at the district level. Employment statistics refer to count of people employed. Row variables are reported in 10,000s. *p<0.1, **p<0.05, ***p<0.01.

For demographic variables, a similar story emerges. Population increases and literacy rates are strongly and positively associated with an increase in VTIs overall (see Tables 7 and 8). Strikingly, an increase of 10,000 marginalized individuals belonging to Scheduled Castes (SC) is correlated with an increase of 4.5 private VTIs. Scheduled Castes are a marginalized and socio-economically backward community in India and the association aligns well with prior work indicating lower caste households make up a sizeable proportion of VTI attendees (Kumar et al., 2019).

Table 7: Demographic determinants of: number of Govt VTIs

	No. of Govt. VTIs					
	(1)	(2)	(3)	(4)	(5)	(6)
Total Population	0.184***					
	(0.042)					
Total Female Population		0.360***				
		(0.089)				
SC Population			0.307*			
			(0.172)			
ST Population				0.342		
				(0.229)		
Total Literates					0.227***	
					(0.044)	
No. of HH						0.633***
						(0.183)
Dep. variable mean	75.49	75.49	75.88	73.77	75.49	75.49
Observations	4020	4020	3997	3830	4020	4020
No. of districts	582	582	582	582	582	582
No. of years	7	7	7	7	7	7
R^2	0.95	0.95	0.95	0.95	0.95	0.95

NOTES—Table presents results from a set of bivariate regressions. Each row displays results from a separate regression. Specification includes district and year fixed effects. Standard errors clustered at the district level. SC and ST Population denote the population for individuals belonging to Scheduled Caste and Scheduled Tribe respectively, both of which are lower caste groups in India. Row variables are reported in 10,000s. *p<0.1, **p<0.05, ***p<0.01.

Table 8: Demographic determinants of: number of Private VTIs

	No. of Private VTIs					
	(1)	(2)	(3)	(4)	(5)	(6)
Total Population	0.923***					
	(0.234)					
Total Female Population		1.920***				
		(0.472)				
SC Population			4.583***			
			(0.779)			
ST Population				0.512		
				(0.508)		
Total Literates					1.057***	
					(0.223)	
No. of HH						3.070***
						(0.782)
Dep. variable mean	50.44	50.44	50.73	51.78	50.44	50.44
Observations	4020	4020	3997	3830	4020	4020
No. of districts	582	582	582	582	582	582
No. of years	7	7	7	7	7	7
R^2	0.67	0.68	0.68	0.64	0.69	0.66

NOTES—Table presents results from a set of bivariate regressions. Each row displays results from a separate regression. Specification includes district and year fixed effects. Standard errors clustered at the district level. SC and ST Population denote the population for individuals belonging to Scheduled Caste and Scheduled Tribe respectively, both of which are lower caste groups in India. Row variables are reported in 10,000s. *p<0.1, **p<0.05, ***p<0.01.

The previous analysis did not account for the correlation between demographic and economic statistics. An increase in population and literacy strongly tracks the increase in total employment. The following set of results shown in Tables 9 & 10 account for this by adding the set of census covariates in the panel regression. I also extend the span of outcomes to include an extensive margin measure for VTI entry (column 2) and focus on both the number of trades and their content (columns 3-5).

Table 9: Govt ITIs only: Determinants of ITIs and trades offered

	(1) No. of Govt. VTIs	(2) Any new VTI?	(3) No. of unique trades	(4) Engineering trades?	(5) Non-engineering trades?
Panel A: Demographics [in 10,000s]					
Total Population	2.332***	-0.011	0.079	0.008**	0.006*
	(0.420)	(0.010)	(0.058)	(0.003)	(0.003)
Total Female Population	-5.515***	0.039*	-0.172	-0.013*	-0.012*
	(0.933)	(0.021)	(0.123)	(0.007)	(0.007)
SC Population	-1.486***	-0.002	-0.038	0.000	0.002
	(0.294)	(0.005)	(0.031)	(0.002)	(0.002)
ST Population	-0.062	-0.006*	-0.018	-0.001*	-0.000
	(0.143)	(0.003)	(0.017)	(0.001)	(0.001)
Total Literates	1.103***	-0.009***	0.095***	-0.001	0.000
	(0.130)	(0.003)	(0.015)	(0.001)	(0.001)
No. of HH	-1.595***	0.002	-0.307***	-0.006**	-0.007**
	(0.370)	(0.008)	(0.050)	(0.003)	(0.003)
Panel B: Employment [in 10,000s]					
Total employment	-0.402	-0.015	-0.163	-0.007	-0.003
	(0.905)	(0.019)	(0.100)	(0.008)	(0.007)
Informal employment	-0.181	-0.002	0.008	0.001	-0.000
	(0.111)	(0.002)	(0.010)	(0.001)	(0.001)
Private employment	-0.023	0.017	0.164*	0.008	0.005
	(0.834)	(0.016)	(0.085)	(0.007)	(0.006)
Manuf. employment	0.382	-0.021	-0.028	-0.004	-0.004
	(0.961)	(0.017)	(0.088)	(0.007)	(0.007)
Services employment	1.029	-0.004	0.009	-0.001	-0.000
	(0.786)	(0.015)	(0.072)	(0.007)	(0.006)
Dep. Var Mean	73.99	0.24	16.83	0.85	0.78
Observations	2043	1569	2043	2043	2043
No. of districts	531	531	531	531	531
No. of years	4	3	4	4	4
R^2	0.96	0.44	0.96	0.76	0.82

NOTES— Each column is a separate regression with the column header as the dependent variable and the rows as set of regressors. Specification includes district and year fixed effects. Standard errors clustered at the district level. Employment statistics refer to count of people employed. SC and ST Population denote the population for individuals belonging to Scheduled Caste and Scheduled Tribe respectively, both of which are lower caste groups in India. *p<0.1, **p<0.05, ***p<0.01

In Table 9 column 2, the likelihood of entry for a government VTI increases by 4 percentage points with an increase in 10,000 women. This result may be picking up a government effort to reduce gender seg-

regation in vocational training and boost female labor force participation. By setting up institutes closer to womens' homes alleviates mobility barriers and loosens norms around womens' work. Overall, with the exception of manufacturing employment (Table 10 column 2), employment variables don't appear to influence VTI entry across the private and public domain. In column 1 of Table 10, I find some evidence that illiteracy and density are associated with decrease in the number of private institutes. This is an amplification of a pattern I find for government VTIs and is perhaps illustrative of the expanding coverage of vocational institutions and migration to urban areas. Remarkably, the association with private VTI entry and SC population is preserved in this set of analysis, making this an interesting line of inquiry for future work.

Table 10: Pvt ITIs only: Determinants of ITIs and trades offered

	(1) No. of Private VTIs	(2) Any new VTI?	(3) No. of unique trades	(4) Engineering trades?	(5) Non-engineering trades?
Panel A: Demographics [in 10,000s]					
Total Population	-2.149	0.016	0.010	0.004	0.002
	(2.180)	(0.013)	(0.066)	(0.005)	(0.005)
Total Female Population	3.602	-0.019	-0.058	-0.001	-0.006
	(4.662)	(0.027)	(0.132)	(0.010)	(0.012)
SC Population	0.438	0.019***	-0.016	0.008**	-0.000
	(1.178)	(0.006)	(0.030)	(0.003)	(0.003)
ST Population	0.282	0.003	-0.003	0.002	0.001
	(0.567)	(0.002)	(0.014)	(0.001)	(0.002)
Total Literates	3.133***	0.003	0.050***	-0.002	0.004***
	(0.606)	(0.003)	(0.011)	(0.001)	(0.001)
No. of HH	-7.126***	-0.032***	-0.059*	-0.003	-0.008*
	(1.656)	(0.008)	(0.034)	(0.005)	(0.005)
Panel B: Employment [in 10,000s]					
Total employment	4.413	-0.025	0.038	0.012	-0.003
	(3.082)	(0.027)	(0.077)	(0.012)	(0.012)
Informal employment	-0.552	-0.002	-0.024*	0.003*	-0.000
	(0.447)	(0.004)	(0.013)	(0.001)	(0.002)
Private employment	-1.478	-0.011	-0.076	-0.027***	-0.006
	(2.813)	(0.022)	(0.062)	(0.007)	(0.008)
Manuf. employment	-4.962	0.047*	0.023	0.012	0.012
	(3.387)	(0.025)	(0.082)	(0.011)	(0.012)
Services employment	-2.562	0.013	0.085	0.000	0.004
	(2.814)	(0.023)	(0.077)	(0.011)	(0.011)
Dep. Var Mean	54.97	0.39	4.88	0.53	0.38
Observations	2043	1569	2043	2043	2043
No. of districts	531	531	531	531	531
No. of years	4	3	4	4	4
R^2	0.69	0.56	0.92	0.77	0.77

NOTES— Each column is a separate regression with the column header as the dependent variable and the rows as set of regressors. Specification includes district and year fixed effects. Standard errors clustered at the district level. Employment statistics refer to count of people employed. SC and ST Population denote the population for individuals belonging to Scheduled Caste and Scheduled Tribe respectively, both of which are lower caste groups in India. *p<0.1, **p<0.05, ***p<0.01

Turning to the number and types of trades, I find that government VTIs exhibit little association with changes in economic and demographic conditions over time. Only increases in private employment and literacy offer a margin for an expansion in the set of trades offered (Table 9, column 3). This pattern is reproduced when restricting the analysis to private VTIs in Table 10. Literacy remains a strong determinant of the range of trades and the association likelihood of engineering and non-engineering trades is strong but not large enough. Disaggregating the categorization of trades is a logical next step that could shed more light on the margins for curriculum change, if at all.

6 Conclusion & Next Steps

6.1 Limitations

The analysis in the paper is not without limitations. Reverse causality is a potential cause for concern in the panel regressions where I regress the number of VTIs on contemporaneous economic and demographic outcomes. It may be that, especially in the initial phase of government led VTI expansion, the entry of vocational institutes spurs responses in employment statistics as the labor force gets skilled. To address this concern, I am working on developing a set of robustness checks where the regressors are lagged by one census year.

Another noteworthy limitation that I allude to in Section 3 is that I don't observe changes in the trade offerings of already established VTIs. This feature, mechanically, limits the change in number of trades offered to either remain the same or increase over time, which might not reflect the true data generating process. I am currently looking into trade affiliation details to extract dates on which certain trades—within a VTI—may have gained or lost affiliation. If feasible, this would be a significant improvement in addressing the issue.

Finally, my analysis may be underpowered to detect correlations, which is another limitation. While, I have more than 500 units and 3 years for every panel regression, there is room to enhance statistical power. This can be done through a few ways: interpolating the economic variables to the census years is a high return option. Relatedly, I can bring the panel to a yearly, rather than at the census year, frequency for more time units to use for analysis.

6.2 Plan ahead

Within the Indian context, this paper is a useful starting point to illustrate the current landscape of a policy-relevant topic: vocational training institutions. Assembling the data to highlight key trends yields some new insights on the supply of VTIs, especially the contrast between privately- and publicly-run institutes. While, there are many potentially interesting strands to explore, I would like to focus on competition between private and public provision of VTIs for extending this project. Given private VTIs are poorer in quality, investigating whether the surge of private VTIs stems from capacity utilization of public VTIs would be first order. If so, figuring out the barriers to expansion and costs associated with provision—maybe as a dimension to compete on—would be the next logical step. The salient quality gradient and whether this is known to students is yet another fruitful area of further inquiry. I hope to bring some survey data to establish the information asymmetries in this market and explore if simple interventions can alleviate the frictions.

References

Andrabi, T., J. Das, and A. I. Khwaja (2015). Delivering education: a pragmatic framework for improving education in low-income countries. In *Handbook of international development and education*, pp. 85–130. Edward Elgar Publishing.

Asher, S., T. Lunt, R. Matsuura, and P. Novosad (2021). Development research at high geographic resolution: an analysis of night-lights, firms, and poverty in india using the shrug open data platform. *The World Bank Economic Review 35*(4), 845–871.

Bank, A. D. (2022). Social inclusion and equity in vocational education and training in india. Technical report, Asian Development Bank. Accessed: 2024-08-24.

Coarasa, J., J. Das, and J. Hammer (2014). Private vs. public. *Finance & Development 51*(4), 35.

Das, J. and J. Hammer (2014). Quality of primary care in low-income countries: facts and economics. *Annu. Rev. Econ. 6*(1), 525–553.

Genicot, G. and D. Ray (2020). Aspirations and economic behavior. *Annual Review of Economics 12*, 715–746.

International Labour Organization (2022). Global employment trends for youth 2022. Technical report, International Labour Organization. Accessed: 2024-08-24.

Jensen, R. (2012). Do labor market opportunities affect young women's work and family decisions? experimental evidence from india. *The Quarterly Journal of Economics 127*(2), 753–792.

Kumar, R., S. Mandava, and V. S. Gopanapalli (2019). Vocational training in india: Determinants of participation and effect on wages. *Empirical Research in Vocational Education and Training 11*(1), 1–17.

Kumar, S. (2022). The skilling imperative in india: The bridge between women and work.

Mehrotra, S., R. Raman, N. Kumra, Kalaiyarasan, and D. Röß (2020). Vocational education and training reform in india. Technical report, Asian Economic Papers.

NITI Aayog, . (2023). Transforming industrial training institutes. Technical report, New Delhi, India.

Prillaman, S. A., R. Pande, V. Singh, and C. T. Moore (2017). What constrains young indian women's labor force participation? evidence from a survey of vocational trainees. Technical report, Working paper.

Rageth, L. and A. Sritharan (2022). Decision-making under information asymmetry. Technical report, Center for Economic Studies (CES). Accessed: 2024-08-24.

UIS, . (2016). Unesco institute for statistics. Technical report.

A Appendix: Additional Tables

A.1 Determinants of number of trades

Table A1: Demographic determinants of: number of trades offered by Govt VTIs in district

	No. of unique trades					
	(1)	(2)	(3)	(4)	(5)	(6)
Total Population	0.009***					
	(0.003)					
Total Female Population		0.017**				
		(0.007)				
SC Population			0.031*			
			(0.019)			
ST Population				-0.020		
				(0.018)		
Total Literates					0.011***	
					(0.004)	
No. of HH						-0.001
						(0.013)
Dep. variable mean	17.41	17.41	17.48	16.94	17.41	17.41
Observations	4020	4020	3997	3830	4020	4020
No. of districts	582	582	582	582	582	582
No. of years	7	7	7	7	7	7
R^2	0.96	0.96	0.96	0.96	0.96	0.96

NOTES—Table presents results from a set of bivariate regressions. Each row displays results from a separate regression. Specification includes district and year fixed effects. Standard errors clustered at the district level. SC and ST Population denote the population for individuals belonging to Scheduled Caste and Scheduled Tribe respectively, both of which are lower caste groups in India. Row variables are reported in 10,000s. *p<0.1, **p<0.05, ***p<0.01.

Table A2: Economic determinants of: number of trades offered by Govt VTIs in district

	No. of unique trades				
	(1)	(2)	(3)	(4)	(5)
Total employment	-0.023				
	(0.015)				
Informal employment		0.016*			
		(0.008)			
Private employment			-0.022		
			(0.015)		
Manuf. employment				-0.078*	
				(0.041)	
Services employment					-0.025
					(0.019)
Dep. variable mean	17.51	17.51	17.51	17.51	17.51
Observations	2271	2271	2271	2271	2271
No. of districts	582	582	582	582	582
No. of years	4	4	4	4	4
R^2	0.96	0.96	0.96	0.96	0.96

NOTES—Table presents results from a set of bivariate regressions. Each row displays results from a separate regression. Specification includes district and year fixed effects. Standard errors clustered at the district level. Employment statistics refer to count of people employed. Row variables are reported in 10,000s. *p<0.1, **p<0.05, ***p<0.01.

Table A3: Demographic determinants of: number of trades offered by Private VTIs in district

	No. of unique trades					
	(1)	(2)	(3)	(4)	(5)	(6)
Total Population	0.014***					
	(0.004)					
Total Female Population		0.030***				
		(0.009)				
SC Population			0.095***			
			(0.016)			
ST Population				-0.007		
				(0.013)		
Total Literates					0.018***	
					(0.004)	
No. of HH						0.060***
						(0.017)
Dep. variable mean	4.78	4.78	4.81	4.77	4.78	4.78
Observations	4020	4020	3997	3830	4020	4020
No. of districts	582	582	582	582	582	582
No. of years	7	7	7	7	7	7
R^2	0.90	0.90	0.90	0.91	0.90	0.90

NOTES—Table presents results from a set of bivariate regressions. Each row displays results from a separate regression. Specification includes district and year fixed effects. Standard errors clustered at the district level. SC and ST Population denote the population for individuals belonging to Scheduled Caste and Scheduled Tribe respectively, both of which are lower caste groups in India. Row variables are reported in 10,000s. *p<0.1, **p<0.05, ***p<0.01.

Table A4: Economic determinants of: number of trades offered by Private VTIs in district

	No. of unique trades				
	(1)	(2)	(3)	(4)	(5)
Total employment	0.061***				
	(0.017)				
Informal employment		-0.027**			
		(0.011)			
Private employment			0.064***		
			(0.018)		
Manuf. employment				0.123**	
				(0.050)	
Services employment					0.090***
					(0.023)
Dep. variable mean	4.82	4.82	4.82	4.82	4.82
Observations	2271	2271	2271	2271	2271
No. of districts	582	582	582	582	582
No. of years	4	4	4	4	4
R^2	0.90	0.90	0.90	0.90	0.90

NOTES—Table presents results from a set of bivariate regressions. Each row displays results from a separate regression. Specification includes district and year fixed effects. Standard errors clustered at the district level. Employment statistics refer to count of people employed. Row variables are reported in 10,000s. *p<0.1, **p<0.05, ***p<0.01.

A.2 Private vs govt VTIs: Robustness to district fixed effects

Table A5: Cross-sectional variation across private- and government-run VTIs

	Government ITIs	Difference	N
	(1)	(2)	(3)
No. of trades offered	8.494	-5.906***	14759
	[6.394]	(0.150)	
Years since establishment	28.033	-11.033***	14211
	[18.874]	(0.608)	
In urban location	0.283	-0.035***	14759
	[0.450]	(0.013)	
Engineering courses offered	0.954	-0.019***	14759
	[0.210]	(0.006)	
Non-engineering courses offered	0.746	-0.411***	14759
	[0.435]	(0.017)	
Standardized VTI quality score	0.000	-0.695***	14517
	[1.000]	(0.045)	

NOTES—Specification includes district fixed effects. Standard Deviation reported in square brackets and standard errors in parentheses. Standard errors clustered at the district level. *p<0.1, **p<0.05, ***p<0.01.

Table A6: Correlates of VTI quality

	Standardized VTI quality score			
	(1)	(2)	(3)	(4)
Years since establishment	0.016***			
	(0.001)			
Private VTI		-0.597***		
		(0.042)		
In urban location			0.066***	
			(0.025)	
No. of trades offered				0.042***
				(0.002)
Dep. variable mean	0.01	0.01	0.01	0.01
Observations	14013	14559	14559	14559
No. of districts	575	578	578	578
No. of years	-	-	-	-
R^2	0.16	0.17	0.12	0.15

NOTES—Table above presents results from bivariate regressions in a cross-sectional dataset. Dependent variable is a standardized version of the overall VTI score, representing quality. Specification includes district fixed effects. Standard Deviation reported in square brackets and standard errors in parentheses. Standard errors clustered at the district level. *p<0.1, **p<0.05, ***p<0.01.

A.3 Panel Regressions: All VTIs

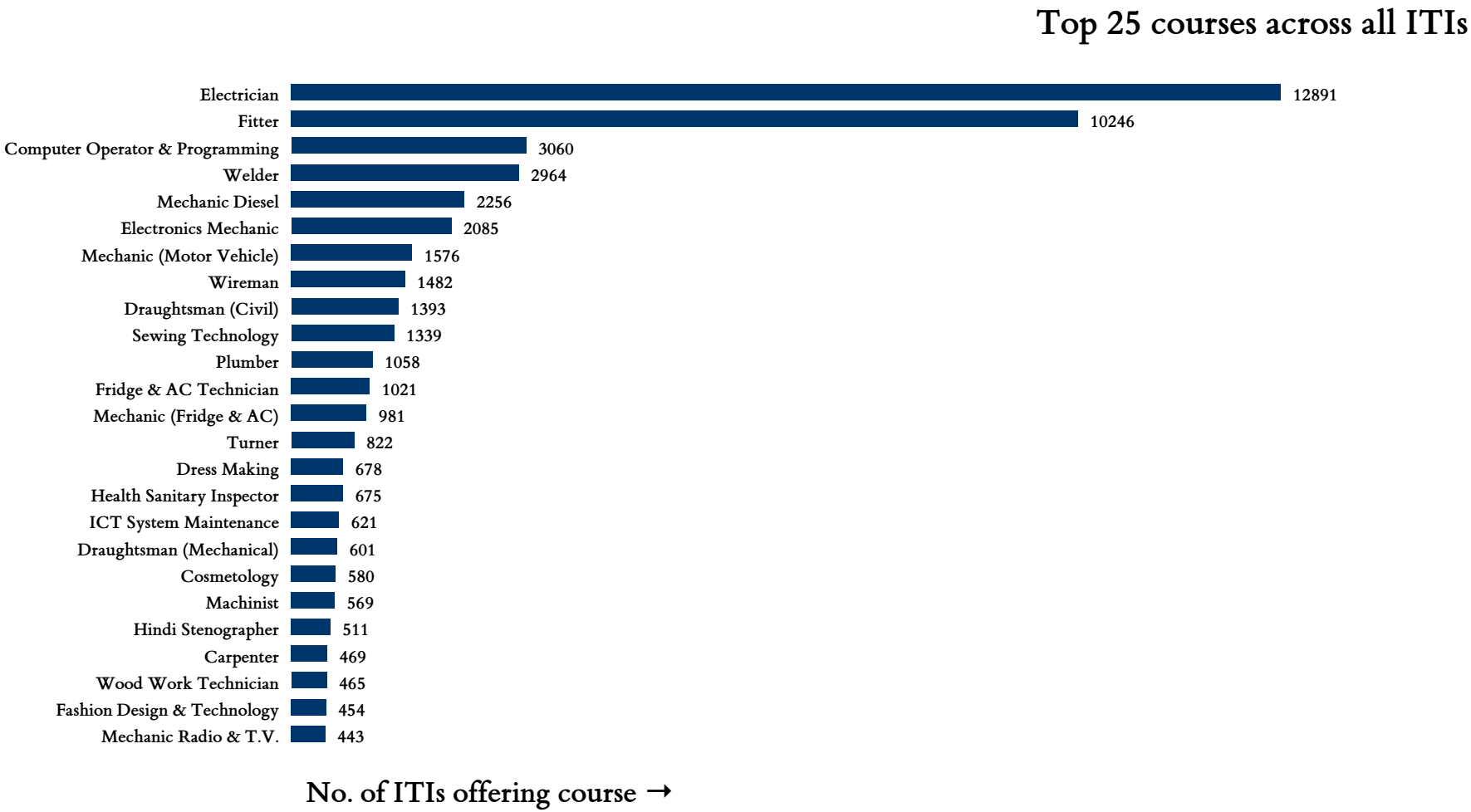
Table A7: Determinants of ITIs and trades offered

	(1) No. of VTIs	(2) Any new VTI?	(3) No. of unique trades	(4) Engineering trades?	(5) Non-engineering trades?
Panel A: Demographics [in 10,000s]					
Total Population	0.183	-0.003	0.017	0.008***	0.008**
	(2.286)	(0.011)	(0.059)	(0.003)	(0.004)
Total Female Population	-1.913	0.021	-0.070	-0.012*	-0.016**
	(4.829)	(0.025)	(0.128)	(0.007)	(0.008)
SC Population	-1.048	0.010*	-0.034	0.001	0.001
	(1.217)	(0.006)	(0.031)	(0.002)	(0.002)
ST Population	0.220	0.000	-0.017	-0.001*	-0.001
	(0.551)	(0.002)	(0.019)	(0.001)	(0.001)
Total Literates	4.236***	-0.001	0.105***	-0.002***	-0.000
	(0.599)	(0.003)	(0.015)	(0.001)	(0.001)
No. of HH	-8.721***	-0.018**	-0.294***	-0.005**	-0.003
	(1.704)	(0.008)	(0.052)	(0.002)	(0.003)
Panel B: Employment [in 10,000s]					
Total employment	4.011	-0.019	-0.037	0.003	0.003
	(3.120)	(0.025)	(0.096)	(0.007)	(0.008)
Informal employment	-0.733	-0.001	-0.001	0.001	0.001
	(0.488)	(0.003)	(0.013)	(0.001)	(0.001)
Private employment	-1.501	-0.020	0.088	-0.003	-0.001
	(2.717)	(0.021)	(0.080)	(0.003)	(0.006)
Manuf. employment	-4.579	0.034	-0.080	-0.002	-0.005
	(3.653)	(0.024)	(0.089)	(0.007)	(0.007)
Services employment	-1.533	0.022	-0.022	-0.001	-0.006
	(2.955)	(0.021)	(0.078)	(0.006)	(0.007)
Dep. Var Mean	128.96	0.52	18.34	0.89	0.82
Observations	2043	1569	2043	2043	2043
No. of districts	531	531	531	531	531
No. of years	4	3	4	4	4
R^2	0.84	0.51	0.97	0.71	0.77

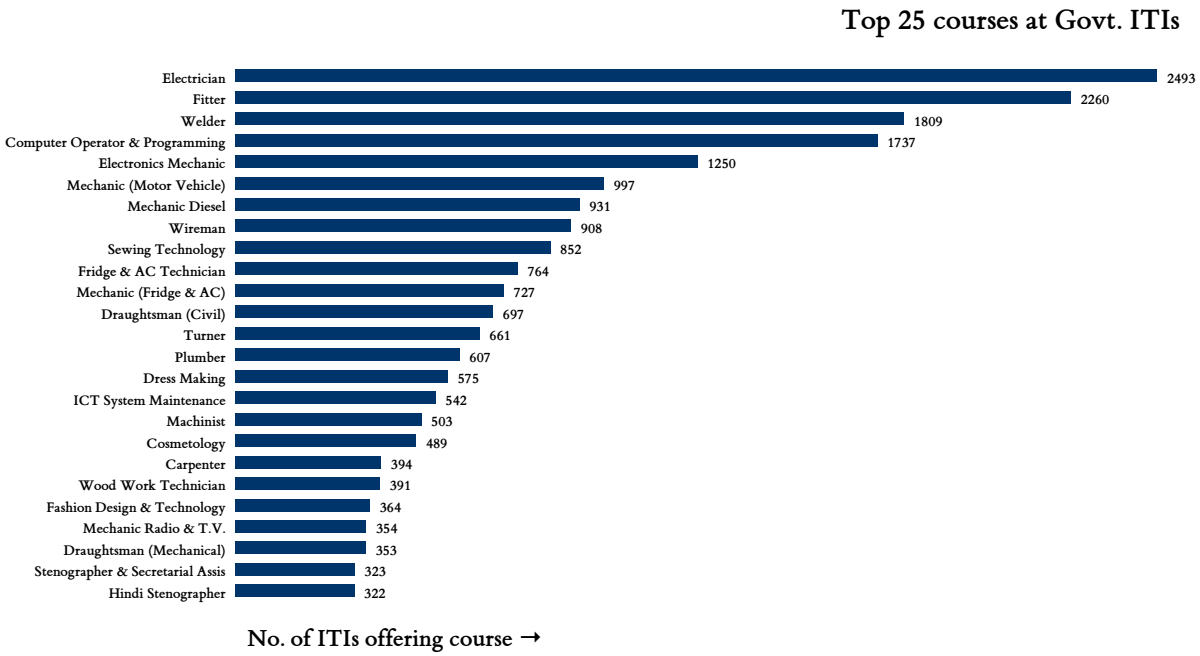
NOTES— Each column is a separate regression with the column header as the dependent variable and the rows as set of regressors. Specification includes district and year fixed effects. Standard errors clustered at the district level. Employment statistics refer to count of people employed. SC and ST Population denote the population for individuals belonging to Scheduled Caste and Scheduled Tribe respectively, both of which are lower caste groups in India. *p<0.1, **p<0.05, ***p<0.01

B Popular Trades

B.1 Overall



B.2 Government VTIs



B.3 Private VTIs