Working Paper no. **146**

Identifying educational needs by rationalising TVET industry clusters in Brazil

Paulina N. Marra

University of Turin, Dipartimento di Economia e Statistica "Cognetti de Martiis", Italy

December, 2015

Laboratorio R. Revelli, *Collegio Carlo Alberto*
Via Real Collegio, 30 - 10024 Moncalieri (TO)

Tel. +39 011 670.50.60 - Fax +39 011 670.50.61
www.laboratoriorevelli.it - labor@laboratoriorevelli.it

Identifying educational needs by rationalising TVET industry clusters in Brazil

Paulina N. Marra[1]

Dipartimento di Economia e Statistica "Cognetti de Martiis", Università degli Studi di Torino, Turin, Italy

Acknowledgements

This research was supported by SENAI with information of its revenues databases and Catalogue of Courses. I thank Márcio Guerra Amorim, Labour Economist, and Luis Eduardo Madeiro Guedes, Statistician, for assistance with the cluster analysis, and also thank Prof. Roberto Leombruni and Prof. Matteo Richiardi, Researchers at University of Turin, who provided insight and expertise that greatly assisted the research, although they may not agree with all of the interpretations and conclusions of this paper.

Abstract

Technical Vocational Educational and Training (TVET) institutions need to reinvent themselves to keep up with market trends, which is why they should assume a more proactive strategy instead of a reactive one towards their Catalogue of Courses. Assuming that TVET institutions are also revenue-maximising institutions, they must provide all of the information that employers and workers need in their materials in a common language and should avoid multiple marketing actions. Therefore, the aim of this research is to present an empirical study to group industrial segments with similar occupational structures and R&D investments to help TVET institutions identify their training needs and tailor their descriptive materials of their marketing and sales teams. Data mining and cluster analysis were used to identify how regular courses at the Brazilian National Service for Industrial Training (SENAI) could be offered within industrial segments in 2011.

JEL Classification: I29, J24, M53, R11

Key Words: TVET, professional occupations, industry clusters, R&D, Catalogue of Courses, labour market

[1] Email: pdossant@unito.it / paulina.marra@gmail.com

1. Introduction

Technical Vocational Educational and Training (TVET) institutions offer different types of training to overcome job mismatch issues by targeting professional occupations with similar professional profiles across industrial sectors, thus enhancing the reach of such training (UNESCO, 2013).

The two main challenges of TVET institutions are to propagate the concept that learning experiences may occur in a variety of different contexts and to provide specific educational requirements that improve the productivity of workers (or at minimum bring them up to an adequate skill level) in specific fields. Lauglo defined education as 'all forms of deliberate intervention designed to bring about learning and training specifically aimed at achieving mastery of performance in specified roles and tasks' (893, 2009).

In addition, as Silva and Hewings (2010) explained, this 'mastery of performance' seen as an increase in productivity influences the manner in which firms select candidates for existing positions. They are more willing to hire individuals that have the correct skillset rather than those who require training and further investment.

This choice is made not only to reduce training costs but also to reduce staff turnover. In other words, the correct amount of targeted training could be provided to the candidate to reduce the risk of competitor 'poaching' of staff. Because well-trained workers perform more efficiently and are, consequently, more productive, human capital is a vital element for competitiveness (Becker, 1994).

Therefore, if the training course is not selected properly because of misleading choices or a lack of recognition of courses within the catalogue, these problems could affect workers' professional life and lead to skill mismatch.

Therefore, establishing TVET's Catalogue of Courses is an important issue for any country, especially for institutions that depend on wholesale revenues. Offering demand-driven and updated training attracts economic investments to a country or a region. Moreover, a qualified labour force means a more productive labour force and reduced training costs for the employer.

The aim of this study is to demonstrate that industrial segments with similar occupational structures and investments in R&D can be grouped to enable TVET institutions to identify the training needs of these segments, offer training along different segments and tailor their descriptive materials to optimise the efforts of the marketing sales force team.

The methodology proposes updating the TVET's Catalogue of Courses by connecting it to professional occupational and economic activity codes. Moreover, data on R&D investments by industrial segment, data mining and cluster analysis techniques may be used to group a 2011 sample of 6,249 industries representing 2,575,342 employers to determine a technology index. This technology index may be used to identify how the regular courses at the Brazilian National Service for Industrial Training (SENAI)[2] can be offered within industrial segments.

The industrial sectors[3] were analysed by professional occupational structure, by workplace accident risk and by technological intensity and, as a result, grouped into nine clusters. In addition, although the industrial segments were expected to be preserved in isolated clusters, given the diversity of this

---

2 Brazilian National Service for Industrial Training (SENAI) is a private institution created in 1942 by the employers' association to supply the Brazilian industry with professional training and qualification technological innovation. SENAI is the largest professional training complex and TVET institution in Latin America, training more than two million Brazilian workers every year (See www.senai.br).
3 Industrial sectors represent the fields of industrial economic activity in which the industry operates. For that classification, we considered just the economic activity of SENAI's contributors.

industrial segment, manufacturing segment needed to be disaggregated to better represent the occupational structure.

2. Method

2.1. Market Segmentation

Market segmentation interferes in not only the way registers are imputed in the database but also how consumers perceive and purchase a catalogue of courses. Nevertheless, the focus of the TVET's district offices were on attempting to satisfy employers' immediate needs, even if neither current technology nor labour market trends were being discussed to create an identity for the TVET's Catalogue of Courses.

This study started in 2011 with the aim to propose a methodology to group TVET's regular courses to better organise the Catalogue of Courses into the marketing material to be offered within industry segments. In addition, several aspects of database mining needed to be ascertained before any attempt was made to model it by occupational structure.

First, the industry behaviour was analysed through the observation of its affiliate industries (e.g. total employees, type of occupations) by economic activity. The focus was on the correlation of industry segments (CNAE[4]) and professional occupations (CBO[5]) (see Figure 1).

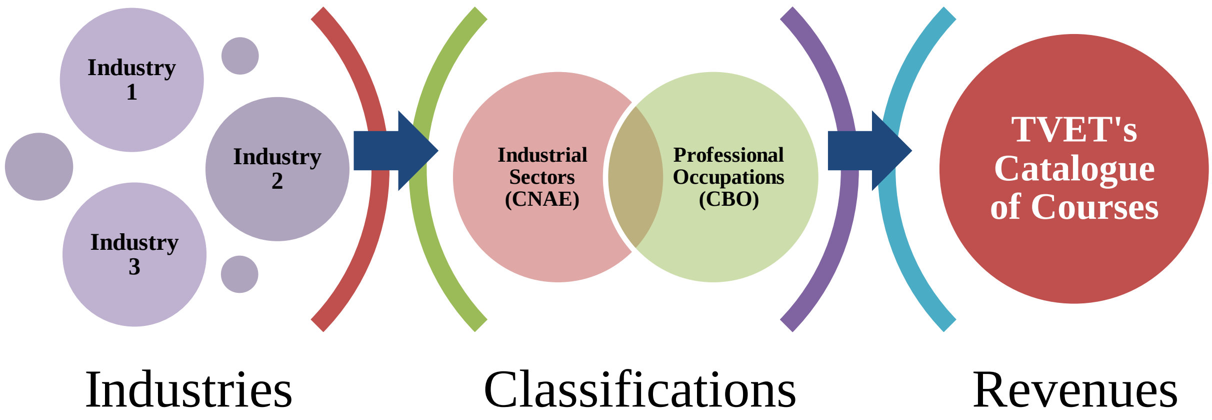

Figure 1–Crossing industrial sectors and professional occupations to establish TVET's catalogue of courses

The second step was to analyse the occupational structure through the 32 industrial sectors[6] in which the TVET courses were segmented. (See table 1)

4 The Brazilian National Classification of Economic Activities (CNAE-2.1) is a standard instrument of classification to identify productive entities in Brazil synched with the ISIC-rev4.
5 The Brazilian Occupation Classification (CBO-2002) describes the occupations and occupation families of the Brazilian labour market and is synched with the ISCO-88.
6 Marra, Paulina N. (2015) Labour Market Segmentation for Industries (work in progress).

Table 1–SENAI's 2012 Industrial Sectors

Industrial Sector–Level 1	**Industrial Sector–Level 2**
Communication and Informatics	Communication; Informatics
Construction	Buildings; Infrastructure; Specialised Services
Manufacturing	Food and Beverages; Cellulose and Paper; Leather and Footwear; Electro electronics; Medical Equipment, Hospital, Optical and Precision; Tobacco; Graphics and Editorial; Jewellery and Gemmology; Wood and Furniture; Metal Mechanics; Non-Metallic Minerals; Other Transport Equipment; Oil, Gas and Fuels; Plastics and Rubber; Miscellaneous Products; Chemicals, Petrochemicals and Pharmaceuticals; Clothing and Textiles; Automotive Vehicles
Railway and Pipeline Transportation	Pipeline Transportation; Rail Transportation
Extractive Industry	Metallic Mineral Extraction; Non-Metallic Mineral Extraction; Extraction Oil, Coal and Natural Gas; Forestry Production
Services of Public Utility	Water and Sewer; Electricity and Gas
Industry Support[7]	

2.2. Technological Intensity

The 'skilled-biased technological change' (Borjas, 2012) can act as either substitutes or complements to labour. When acting as a substitute, it reduces the amount of employees hired, shifting the demand curve to the left (↓wages; ↓quantities of labour). The 'skilled-biased technological change' may increase the demand for certain types of labour (skilled workers who adapted to changes in technology) when acting as a complement to labour, shifting the demand curve to the right (↑wages; ↑quantities of labour).

Technology can definitely increase wage inequality, especially when a new technology is introduced (OpenStax College, 2014). Workers need time to adapt to new technologies; therefore, during this period, highly skilled workers will have raised wages and the quantities of labour, affecting low-skilled workers who will have reduced wages and the quantities of labour until the technology is widespread. Then, the market again reaches equilibrium (see Figure 2).

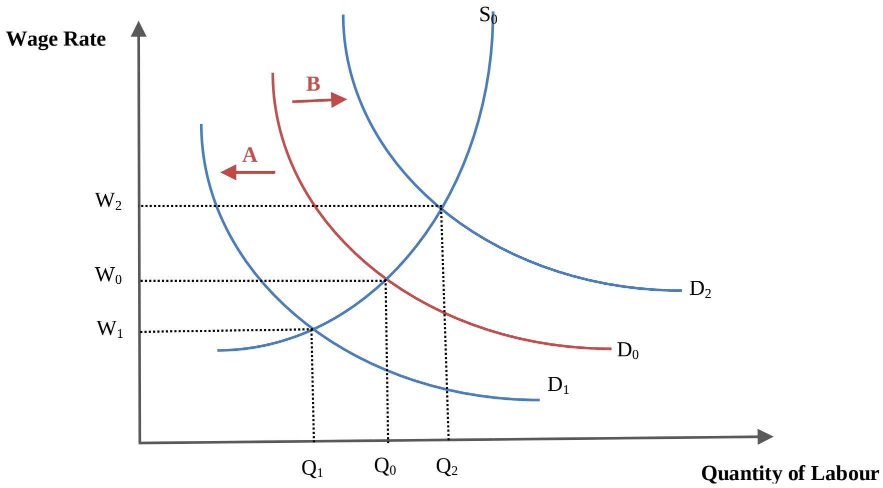

Figure 2–Impact of Technological Change on Supply and Demand Curves

7 Industry support is a non-industrial but industry-related economic activity, such as offices, warehouses, and factory stores, affiliated with industry headquarters.

Therefore, because technology trends need to be inserted quickly into the content of courses content, are reflected in the occupational structure, and have a significant effect on choices of training, they need to be reflected in the marketing materials, thus also influencing purchases.

Thus, identifying the variable that could represent changes in technology in the model, the third step right after defining the market segmentation was crucial for the cluster analysis. Higher investments in R&D require more skilled occupations for the specific economic activity. However, how should the level of technology be measured on the supply side and the demand side?

The R&D intensity index is the most important indicator used by the OECD to classify industries according to their R&D investments. These indexes are extremely important for understanding market trends and future investments in technology.

In developed countries, technological intensity describes the speed of the international technological frontier in general and represents the overall dynamics of that frontier. At this frontier, each country specialises in a small number of industries, which describes the efforts towards the transfer process.

In 2003, the OECD STI Scoreboard (OECD, 2003) presented a study grouping industries by technological intensity to develop a classification on the basis of expenditures in the R&D industries of member countries[8]. This study proposes a sectorial grouping on the basis of the ranking of the mentioned variable by industrial activity. The categories were obtained by dividing the ranking into 'quartiles', i.e. four equal parts (high, medium-high, medium-low and low) (see Table 2).

Table 2–R&D Intensity (OECD)

High Technological Intensity	**Medium-High Technological Intensity**	**Medium-Low Technological Intensity**	**Low Technological Intensity**
Aerospace; Computers; Electronics; Instruments; Pharmaceutical; Telecommunications	Automotive vehicles; Chemical (except pharmaceuticals); Electric Materials; Machinery and equipment; Rail transportation; Transport equipment	Metal Products; Metallurgy; Other non-metallic products; Refined petroleum products and nuclear fuel; Rubber and plastic; Shipbuilding	Cellulose and paper; Editorial and Graphic; Food, beverages and tobacco; Leather and footwear; Other sectors; Recycling; Textiles; Wood

Through PINTEC's[9] results, Furtado and Quadros (Furtado & Carvalho, 2005) proposed the use of the OECD R&D technological intensity index and classification methodology for the Brazilian industry. Then, Franco et al. (Franco, Carvalho, & Carvalho, 2006) propose a method to construct a sectorial aggregation on the basis of indicators of effort and innovative activity. This method relativised information from PINTEC, such as expenditure and allocated personnel in R&D, and the amount of innovative firms.

In 2007, Furtado et al. (Furtado, Quadros, & Domingues, 2007) discussed the advantages of using PINTEC as a parameter to depict the technological and innovative industrial structure on the basis of the 'quartiles' proposed by the OECD. This structure proposes a new arrangement of industrial sectors to analyse the intensity of R&D. The index is calculated as the ratio of R&D spending of a particular

8 A.12.1. "R&D in non-OECD economies" pp. 66–67; D.6. "Technology and knowledge-intensive industries" pp. 140–141; D.9.3. "Revealed comparative advantage by technological intensity" pp. 150–151; Annex I "Classification of manufacturing industries based on technology" pp. 155–157

9 Brazilian Industry Technological Innovation Research (PINTEC) is a survey performed by the Brazilian Institute of Statistics (IBGE) that aims to build national and sectorial indexes comparable across countries. It focus factors that could influence innovation behaviour though strategies, efforts, incentives, barriers and outcomes of innovation.

company to its sales or added value. This indicator is defined as the Technological Intensity index[10](see Table 3).

Table 3–R&D Intensity (PINTEC)

Technological Intensity	PINTEC 2005	Intensity of R&D
High	Manufacture of other transport equipment	3.22
	Activities of computer and related services	2.33
	Instrumentation and industrial automation	2.26
	Office machines and computer equipment	1.48
	Manufacture of machinery, appliances and equipment	1.29
	Vehicles, trailers and wagons	1.25
Medium-High	Electronics and appliances for telecommunications	1.10
	Petroleum refining and alcohol	0.77
	Manufacture of machinery and equipment	0.55
	Manufacture of chemicals	0.55
	Telecommunications	0.52
	Fabrication of furniture and miscellaneous industries	0.47
Medium-Low	Manufacture of rubber and plastic	0.42
	Manufacture of non-metallic minerals	0.36
	Leather, footwear and leather goods manufacturing	0.34
	Manufacture of tobacco products	0.23
	Manufacture of cellulose, paper and paper products	0.23
	Manufacture of articles of clothing and accessories	0.22
	Manufacture of textiles	0.22
Low	Manufacture of metal products	0.21
	Extractive industries	0.20
	Metallurgy	0.18
	Manufacture of wood	0.13
	Manufacture of food and beverages	0.13
	Editorial, graphics and reproduction of recordings	0.08

When applying the intensity of the R&D index to the 32 industrial sectors (level 2) (which means assigning each industrial sector to a proper technological intensity 'quartile') the sectors could be easily distinguished using the potential risk of 'skilled-biassed technological changed' on the supply side (see Table 4).

10 The Technological Intensity index measures the effort in innovation and the technological change from the relationship of costs in research activities and technological development and the added value of these amounts calculated using the PINTEC results.

Table 4–Industrial Sectors Linked to Technological Intensity (PINTEC, 2005)

Industrial Sector Level 1	***Industrial Sector Level 2***	***Industrial Sector Level 3***	***Technological Intensity***
Communication and Informatics	Communication	Film	Medium-High
		Post Office	Medium-High
		Wire Telecommunications	Medium-High
		Wireless Telecommunications	Medium-High
		Data transmission	Medium-High
	Informatics	Hardware	Low
		Software	High
Construction	Buildings		Low
	Infrastructure		Low
	Specialised Services	Demolition and site preparation	Low
		Electrical, hydraulic and other	Low
		Finishing works	Low
		Other specialised services	Low
Manufacturing	Food and Beverages	Food	Low
		Beverages	Low
		Bakery	Low
	Cellulose and Paper	Cellulose	Medium-Low
		Paper	Medium-Low
	Leather and Footwear	Leather Goods and Travel Goods	Low
		Footwear	Low
		Preparation of Leather	Medium-Low
	Electro electronics	Electric	Medium-High
		Electronics	Medium-High
	Medical Equipment, Hospital, Optical and Precision		
	Tobacco		Medium-Low
	Graphics and Publishing	Graphics	Low
		Publishing	Low
	Jewellery and Gemmology		Medium-High
	Wood and Furniture	Wood	Low
		Furniture	Medium-High
	Metal Mechanics	Mechanics	High/Medium-High/Low
		Metallurgy	Low
		Mechanics Maintenance	Low
	Non-Metallic Minerals		Medium-Low
	Other Transport Equipment	Aircrafts	High
		Vessels	High
		Others	High
		Rail Vehicles	High
		Military Vehicles	High
	Oil, Gas and Fuels	Fuels	Medium-High
		Oil and Gas	Medium-High
	Plastics and Rubber	Plastics	Medium-Low
		Rubber	Medium-High
	Miscellaneous Products		
		Pharmaceuticals	Medium-High

	Chemicals, Petrochemicals and Pharmaceuticals	Petrochemicals	Medium-High Medium-Low
		Cleaning Products and Cosmetics	Medium-High
		Chemicals	Medium-High
	Clothing and Textiles	Clothing	Medium-Low
		Textiles	Medium-Low
	Automotive Vehicles	Automotive	High
		Automotive Maintenance	High
		Motorcycles	High
Railway and pipeline transportation	Pipeline Transportation		Low
	Rail Transportation		Low
Extractive Industry	Metallic Mineral Extraction		Low
	Non-Metallic Mineral Extraction		Low
	Extraction Oil, Coal and Natural Gas		Low
	Forestry Production		Low
Services of public utility	Water and Sewer	Water	Low
		Collection and Treatment of Waste	Low
		Sewer	Low
	Electricity and Gas	Electricity	Low
		Gas	Low
		Other Utilities	Low
Industry Support			Low

Nevertheless, because the information on R&D investments came from PINTEC, which had information available only for the manufacturing, extraction and services sectors, consultations with SENAI's technical and macroeconomists experts were conducted. These consultations led to an established set of rules to classify the industrial sector with the proper technological intensity category to overcome this lack of information.

2.3. Sample, Variables and Data Mining

Although the focus of this study is to group data with similar behaviour, it is important to highlight that many adjustments were made before the cluster analysis began. Nevertheless, despite the risks of incompleteness and unreliability, administrative data are still the most robust source of information.

Administrative data need to pass through constant maintenance; therefore, we can benefit the most from data mining, data imputation and micro simulation to overcome the precariousness of the data, such as simulating excess of missing information cycles to extrapolate and validate the model.

Analysing the entire set of Brazilian industries—more than 500,000 firms—was not feasible. Thus, a sample of the 250 largest Brazilian industry headquarters was selected (6,249 affiliate firms and 2,575,342 employers) to further apply and adapt the methodology to the entire set of industries. These firms were selected from a wide range of economic sectors and professional occupations.

The main database used was RAIS[11], which tracks company size[12], their geographic location and economic activity codes, the number of employees, their salaries, occupations and educational level.

Moreover, RAIS could also be linked to SENAI's revenue database, which has information on the courses offered, students enrolled and other factors. Using these administrative databases in a three-dimensional approach could enable analysis to be carried out after data mining adjustments.

Therefore, the following step was undertaken to define the variables that would contribute to building the model and starting the data mining process. Because the principal database had a limited range of variables, these variables were chosen arbitrarily and divided into four groups: economic activity sectors, economic situation (technological intensity level, payroll tax), workforce situation (occupational codes, education) and the workplace accident risk index (RAT[13]).

Data mining was used to find weak relationships between the data in large, noisy and messy transactional databases by using data cleansing and data imputation tools. First, Cartesian rules were adopted for the reasonable value imputation, where the mean value of 7 digits CNAE 2.1, was used to fill in missing information. The data were also adjusted for variables that had missing information on total employees, payroll tax, workplace accident risk and technological intensity.

The main effort in this step was to treat information from classified companies by preserving the criteria from the original ranking; thus, the codes for the economic activity sector serve as the primary key to gathering the information from all of these databases.

The last group of variables to be analysed was related to the workforce situation and uses data on employment contracts provided by RAIS. However, the use of this database faces two limitations: one refers to the characteristics of the classification and the other to the characteristics of the register.

The information at RAIS is presented according to the occupational code family; therefore, if each code were analysed, it would generate a complex database. Because this study is a first attempt to model the Catalogue of Courses—focused on general training—the solution was to choose variables among those for workforce situation and classify occupation groups by type, type of education, qualification level and educational level.

However, the main limitation was that the employment register database was incomplete or needed to be corrected, even though such information is compulsory and was supposed to be binding on administrative records.

Companies usually concentrate most of the administrative functions at headquarters, such as the employees' register. In this case, most of the missing information was concentrated on records related to the employee's professional occupation. Hence, missing data could have a significant impact on data analysis because they affect the statistical significance standard analysis techniques.

Several imputation data techniques exist. This study used the reasonable value imputation and multiple imputation (Rubin, 2004), which calls for assigning data on the basis of reasonable heuristic rules or a set of rules. For these imputations, one or multiple variables help predict the values that are probable on the basis of available information.

Therefore, multiple imputations simulate the process of generating data and the uncertainty associated with the data's probability distribution parameters, thus avoiding biases. Using multiple imputations

11 Brazilian Annual Register of Social Information (RAIS) is part of the Brazilian National Account System and provides the Brazilian government with information regarding the work activity for labour statistics.
12 A Brazilian business classification that groups firms by number of employees (size)
13 Risk of accidents at work (RAT) is an index of workplace accident risk according to the Social Security classification and is related to the economic activity that the worker is carrying out.

was not yet feasible because of the sample and the need to previously understand the behaviour of the missing data before making assumptions.

In this specific case, a set of Cartesian rules was used on the most detailed parameter possible: the CNAE coding level 5, 7 digits. Thus, an average value to the missing information groups was attributed according to the register coding. The variables that passed through the reasonable value imputation process were data on total employees, payroll value, working accident risk and technological intensity and data that refer to companies' workforce (professional occupations).

After the data mining process, the Catalogue of Courses' strategy was evaluated by analysing industries' professional occupations structure in each industrial sector and whether they are affected differently by investments in R&D, company size, quantity of employees, payroll taxes and so on.

The industrial sectors were then analysed using clusters to verify whether the courses offered are different for each sector or whether they could be grouped to present a more cohesive and convergent catalogue to satisfy industry needs.

2.4. Cluster Analysis

As Nisbet et al. (Nisbet, Elder, & Miner, 2009) discussed, cluster analysis is the best method for pattern recognition. Therefore, in this study, data mining algorithms use the association of rules to detect relationships between specific values of categorical variables in the database, assigning those with similar patterns to the same clusters.

The method applied was the TwoStep Cluster Analysis (SPSS, 2001) using Akaike's Information Criterion (AIC). In this method, the algorithm automatically selects the optimal number of clusters using the AIC criterion. The following assumptions were made: the likelihood distance measure assumes that the variables in the cluster model are independent, each continuous variable is assumed to have a normal (Gaussian) distribution and each categorical variable has a multinomial distribution.

The variables used for the cluster analysis were total employees, payroll value, working accident risk index, technological intensity index and data that refer to companies' workforce (professional occupational structure according to type of occupation, type of education, qualification level and educational level).

In addition, despite the fact that some very suitable variables for the model's consistency existed, they were not considered because of a lack of reliable information. Revenue is an example and is an important variable for understanding the profitability and competitiveness of a firm. However, because only one part of the database had this information, a categorised variable as a proxy could compromise the entire model and was excluded.

Therefore, after performing a correlogram analysis, the payroll variable was dropped from the model because it was intrinsically correlated with total employees, probably because of the particularities of the sample (250 largest companies).

The cluster was then grouped in general using the variable industrial sector level 1—except for the manufacturing sector. Using this variable was necessary because the variable industrial sector level 2 had quite different results within its professional occupational structure (workforce situation).

The empirical testing indicates that the procedure is fairly robust to violations of both the independence and the distributional assumptions. The model was built using the variables according to their discriminatory capacity, which was verified from the distribution of data.

3. Results

3.1. Occupational Structure Behaviour

The behaviour analysis begins with a descriptive analysis of the professional occupational structure to identify similar patterns, i.e. behaviours, among the industrial sectors. For this analysis, the segment industry support is exempt from the comments because of the managerial characteristics of the occupations and because it is not a target of this study.

However, before the descriptive analysis, it must be stated that the variable type of education informs the type of formation that an occupation requires. For this requirement, qualification up to 200 h may be interpreted as specific training, qualification over 200 h as general training and technical education as secondary education. In addition, the educational level required by the occupation is based on the same division used by the International Standard Classification of Education (ISCED).

In addition, those variables had missing values issues that prevented the cluster analysis from proceeding. Therefore, data imputation methods were used to equalise that matter and produce significant statistics.

The reasonable value imputation process was quite effective despite the significant percentage of missing values, at 16%, for total employment and payroll tax. For technological intensity and workplace accident risk, the percentage of missing values was very small (0.5%) because they were linked through the economic activity code. After adjustments were made, the number of missing values declined to nearly zero.

For the workforce variables (professional occupational type, professional occupation by type of education, professional occupation by qualification level, professional occupation by educational level), the multiple imputation process was adopted given database limitations. The process was quite effective despite the high percentage of missing values at 38%. Hence, after some adjustments were performed, the percentage of missing values declined to nearly zero.

Regarding the analysis of occupational structure, manufacturing employed the highest percentage of workers in industrial occupations (68.9%), whereas communications and informatics employed the lowest percentage of workers in industrial occupations (30.3%) (Table 5).

Table 5–Professional Occupational Structure according to Type of Occupation

Industrial Sector	**Professional Occupational Type**	
	Industrial	**Non Industrial**
Communication and Informatics	30.3%	69.7%
Construction	52.1%	47.9%
Manufacturing	68.9%	31.1%
Extractive industry	67.3%	32.7%
Railway and pipeline transportation	58.7%	41.3%
Services of public utility	62.7%	37.3%
Industry Support	47.0%	53.0%
Total %	58.5%	41.5%

Despite the high average percentage of non-classified occupations (16.2%), identifying different behaviour through type of education is possible. Communication and informatics employed the highest percentage of workers with a higher education (26.2%), particularly in the Informatics segment. In the other hand, railway and pipelines (1.6%) and public utility services (2.0%) employed the lowest percentages of workers with a higher education. Moreover, the sectors that employed the highest percentages workers are extractive industries in technical education (24.6%), railway and pipelines in qualification over 200 h (37.9%) and manufacturing in qualification up to 200 h (41.5%) (Table 6).

Table 6–Professional Occupational Structure according to Type of Education

Industrial Sector	**Professional Occupation by Type of Education**				
	Qualification up to 200 h	**Qualification over 200 h**	**Technical Education**	**Higher Education**	**Unclassified**
Communication and Informatics	31.1%	10.5%	17.4%	26.2%	14.8%
Construction	31.2%	20.1%	8.9%	22.8%	16.9%
Manufacturing	41.5%	21.8%	10.8%	10.8%	15.1%
Extractive industry	24.4%	22.4%	24.6%	12.1%	16.5%
Railway and pipeline transportation	14.6%	37.9%	19.0%	1.6%	26.9%
Services of public utility	35.2%	22.0%	16.3%	2.0%	24.5%
Industry Support	47.4%	7.9%	9.0%	14.1%	21.6%
Total %	36.3%	19.9%	12.4%	15.2%	16.2%

Communication and informatics employed the highest percentage of workers with low qualifications (56.9%), railway and pipelines employed the lowest percentage of workers with low qualifications (14.7%), extractive industries employed the highest percentage of workers with high qualifications (25.1%) and construction employed the lowest percentage of workers with high qualifications (9.5%) (Table 7).

Table 7–Professional Occupational Structure according to Qualification Level

Industrial Sector	**Professional Occupation by Qualification Level**		
	Low	**Medium**	**High**
Communication and Informatics	56.9%	25.4%	17.6%
Construction	52.4%	38.1%	9.5%
Manufacturing	52.0%	37.4%	10.6%
Extractive industry	27.5%	47.4%	25.1%
Railway and pipeline transportation	14.7%	75.0%	10.3%
Services of public utility	36.8%	46.8%	16.4%
Industry Support	72.1%	20.1%	7.8%
Total %	50.5%	37.1%	12.3%

Extractive industries employed the highest percentage of workers with tertiary educations (26.3%), railway and pipelines employed the highest percentage of workers with a secondary education (84.2%) and construction employed the highest percentage of workers with a primary education (58.1%) (Table 8).

Table 8–Professional Occupational Structure according to Educational Level

Industrial Sector Level 1	**Professional Occupation by Educational Level**			
	Primary education	**Secondary education**	**Tertiary Education**	**Unclassified**
Communication and Informatics	35.5%	46.6%	17.9%	-
Construction	58.1%	32.0%	9.6%	0.3%
Manufacturing	49.4%	39.4%	10.8%	0.3%
Extractive industry	19.9%	52.7%	26.3%	1.0%
Railway and pipeline transportation	4.5%	84.2%	11.2%	0.2%
Services of public utility	25.1%	58.0%	16.8%	0.1%
Industry Support	61.8%	29.4%	7.7%	1.0%
Total %	45.8%	41.3%	12.6%	0.3%

3.2. Industrial Clusters

The descriptive analysis enables identification of a pattern among the industrial sectors. However, determining the intensity and type of relationship that such variables have is impossible. Therefore, a cluster analysis was carried out to group the 32 industrial sectors by tailoring TVET's Catalogue of Courses and marketing materials.

The first step was to identify how the variables would be inserted into the cluster analysis. Because economic activity is the main variable of this model, the analysis needed to be conducted at a more disaggregated level (industrial sector level 2) particularly for manufacturing because it comprised approximately 50% of the database and biassed the clustering results when a first attempt was made to consider only industrial sector level 1.

An attempt was made to use an even more disaggregated level of information for economic activity—industrial sectors level 3. However, because of the sample's size and characteristics, the results were not statistically significant. Therefore, for this study, only industrial sector level 2 is considered.

After applying the TwoStep Cluster method, we verified that, in addition to manufacturing, every industrial sector constituted a single cluster, thus preserving its characteristics. Therefore, the hypothesis that states that distinguishing industrial sectors is possible through occupation structure is supported.

This analysis was performed by cluster through the column percentage and by category through the line percentage and aimed to present the usability of variables in the segmentation model. In other words, a more dispersed or less concentrated distribution resulted in a less discriminatory variable (Table 9).

Table 9–Cluster Analysis by Industrial Sector

	Industrial Sector	Cluster 1			Cluster 2			Cluster 3			Cluster 4			Cluster 5		
		N	%Lin	%Col	N	%Lin	%Col	N	%Lin	%Col	N	%Lin	%Col	N	%Lin	%Col
Level 1	Communications and Informatics				2,480	95.0	49.0				130	4.9	22.7	2	0.1	0.9
	Construction	46	17.3	12.4	161	60.7	3.2				16	6.0	2.8	43	16.0	16.2
	Manufacturing	184	9.6	49.5	1,464	76.2	28.9	83	4.4	33.3	92	4.8	16.2	95	5.0	36.2
	Extractive Industry	94	22.3	25.4	225	53.1	4.5	17	4.2	6.9	45	10.5	7.9	42	9.9	16.0
	Railway and pipeline transportation	7	9.3	1.9	24	31.0	0.5	4	4.6	1.4	6	8.2	1.1	10	47.0	13.6
	Services of public utility	38	22.3	8.8	78	53.2	1.5	4	2.8	1.6	5	3.5	0.9	27	18.2	10.2
	Industry Support	8	0.8	2.0	626	58.5	12.4	142	13.2	56.7	277	25.9	48.5	18	1.7	6.9
Level 2–Manufacturing	Food and Beverages	869	84.8	100.0	55	5.3	24.2	48	4.7	20.8	54	5.2	21.7			
	Cellulose and Paper				32	100.0	14.3									
	Leather and Footwear				12	48.9	5.3				12	51.1	5.0			
	Electro Electronics										1	3.6	0.3	22	96.4	8.8
	Medical Equipment, Hospital, Optical and Precision				-	7.1	0.1	1	14.3	0.2				3	78.6	1.2
	Tobacco													43	100.0	17.5
	Graphics and Publishing				-	1.8	0.1	15	98.2	6.3						
	Wood and Furniture				2	53.3	1.0				2	46.7	0.8			
	Metal Mechanics				96	54.8	42.6	16	9.2	7.0	52	29.8	21.1	11	6.2	4.4
	Non-Metallic Minerals										69	100.0	28.0			
	Other Transport Equipment										-	20.0	0.2	2	80.0	0.9
	Oil, Gas and Fuels				1	1.8	0.4				3	7.8	1.4	40	90.4	16.3
	Plastics and Rubber				24	100.0	10.9									
	Miscellaneous Products							24	61.8	10.4	12	29.9	4.7	3	8.3	1.3
	Chemicals, Petrochemicals and Pharmaceuticals							127	73.3	55.2				46	26.7	18.6
	Clothing and Textiles				1	3.2	0.6				41	96.8	16.5			
	Automotive Vehicles				1	1.7	0.6				-	0.7	0.2	77	97.6	31.0

As a result, the cluster analysis grouped the industrial sectors into nine clusters (Table 10) and the diverse composition of manufacturing was split into five different clusters. Clusters numbered 2, 3 and 4 are industrial sectors well characterised by occupational structure. However, clusters numbered 5—railway and pipelines, services of public utility and industry support—is a combination of sectors.

Nevertheless, the necessity to analyse more disaggregate information becomes more evident when clusters numbered 6, 8 and 9 are analysed. Those clusters share industrial sector level 2—automotive vehicles, metal mechanics and oil, gas and fuels. Therefore, overlap seems to exist in those clusters, which does not occur in reality because the code assignment is done in industrial sectors level 3.

These results should guide the marketing material for regular courses (general training) or the Catalogue of Courses instead of the entire catalogue. The Catalogue may be disaggregated into industrial sector clusters to facilitate the TVET's sales agent approach to firms offering similar professional profiles, which enhance course revenues.

Table 10–Industrial Sector Clusters

Cluster	Industrial Sector–Level 1	Industrial Sector–Level 2
1	Manufacturing	Food and Beverages
2	Communication and Informatics	Communication, Informatics
3	Construction	Buildings, Infrastructure, Specialised Services
4	Extractive Industry	Metallic Mineral Extraction, Non-Metallic Mineral Extraction, Extraction Oil, Coal and Natural Gas, Forestry Production
5	Railway and pipeline transportation, Services of public utility, Industry Support	Pipeline Transportation, Rail Transportation, Water and Sewer, Electricity and Gas, Industry Support
6	Manufacturing	Automotive Vehicles; Cellulose and Paper; Clothing and Textiles; Food and Beverages; Graphics and Publishing; Leather and Footwear; Medical Equipment, Hospital, Optical and Precision; Metal mechanics; Oil, Gas and Fuels; Plastics and Rubber; Wood and Furniture
7	Manufacturing	Chemicals, Petrochemicals and Pharmaceuticals; Food and Beverages; Graphics and Publishing; Medical Equipment, Hospital, Optical and Precision; Metal mechanics; Miscellaneous Products
8	Manufacturing	Automotive Vehicles; Clothing and Textiles; Electro electronics; Food and Beverages; Leather and Footwear; Metal mechanics; Miscellaneous Products; Non-Metallic Minerals; Oil, Gas and Fuels; Other Transport Equipment; Wood and Furniture
9	Manufacturing	Automotive Vehicles; Chemicals, Petrochemicals and Pharmaceuticals; Electro electronics; Medical Equipment, Hospital, Optical and Precision; Metal mechanics; Miscellaneous Products; Oil, Gas and Fuels; Other Transport Equipment; Tobacco

4. Discussion and Recommendations

The effect of skilling on the labour market provides reassurances of how keeping up with technology trends is crucial to meeting the industry's demands and students' expectations (Figure 2). TVET institutions play a distinctive role in this matter.

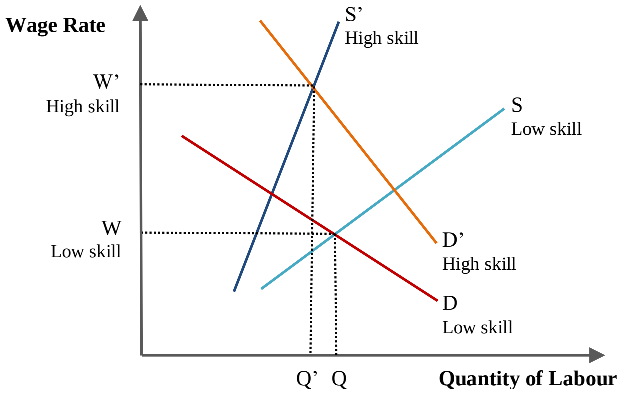

Figure 2–Impact of Skilled Labour in Supply and Demand Curves

For TVET policies to succeed, the most important step is to determine how many people will be trained in different fields and, within each field, the specific and general skills that should be taught (OECD, 2009). However, student preferences and employer needs are key determinants to establishing the Catalogue of Courses. Moreover, better determining product targets makes consumers more inclined to enrol in those courses because the initiative[14] is easier to recognise.

Although student and employer preferences may overlap, some differences will always exist. Moreover, TVET institutions should decide on how many of these differences will be addressed to satisfy students' and employers' needs without compromising the budget or the quality.

Finding the balance between providing students with their preferred training and providing employers with trained workers according to their requirements will increase the probability of the success of a TVET initiative. Such a balance will consistently provide a good match between the skills acquired in education and on-the-job training and the labour market requirements to promote strong and inclusive growth.

Consequently, these positive effects will only be realised if TVET institutions consolidate courses and generate a much higher level of specificity between professional occupations and industrial sectors to meet the demand for skilled workers. This demand can be met by predicting an equivalent proportion of skilled professionals needed by the industry in the future, which is the main goal of TVET policies.

Neugart and Schömann (2003) stated that creating reliable forecasting models is challenging because demand for skills depends on numerous factors, many of which are uncertain and difficult to predict, such as technological progress, global economic conditions and government policies. Nevertheless, because the data are not always available and marketing strategies are much more reactive than proactive, TVET institutions are always one step behind.

However, even if further studies help distinguish the specific level 3 industrial sectors needed in each cluster, the results presented in this study can be applied to general training in TVET institutions.

14 Marra, Paulina N. (2015) The use of taxonomy on vocational education to identify workers' multiple skills and measure mismatch (work in progress)

5. Conclusion

The aim of this study was to identify similar patterns in professional occupations to offer general training within industrial sectors, optimising revenues from the Catalogue of Courses. This aim was accomplished by identifying patterns using professional occupational structure and investments in technology.

As expected, the results presented a behavioural analysis using simple information on investments in R&D (technological intensity), industrial sectors, workplace accident risk and workforce situation, regardless of labour market forecasts or information on market trends.

As presented, investment in R&D is the variable that most influences an employer's training needs, thereby influencing labour market dynamics and hopefully avoiding shortages in future skills. Therefore, an updated catalogue is fundamental to influencing a student's preferences: if the catalogue is consistent with market trends and employers' needs, shortages will decrease in theory.

This study aimed to understand the global behaviour of offering general training through the professional occupational structure of industrial sectors. The industrial sectors were analysed by professional occupational structure, workplace accident risk and technological intensity, and the sectors were then grouped into nine clusters. Although proving the suitability of this segmentation was not the main objective of this study, the industrial sectors were expected to be preserved into isolated clusters.

Moreover, except for manufacturing, which was disaggregated into level 2, the results did not present much of a change from what was being performed to date with the Catalogue of Courses. Although manufacturing represents more than 50% of the SENAI's enrolments, it achieved significant results and can be used for general training to guide the sales marketing team until further studies are carried out using a more disaggregated industrial sector level 3 with a different or larger sample.

6. References

Becker, Gary S. 1994. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. 3rd ed. Chicago: University of Chicago Press.

Borjas, George J. 2012. *Labor Economics*. 6th ed. McGraw-Hill Education.

Franco, Eliane, Flávia Carvalho, and Sílvia Carvalho. 2006. "Como Construir Grupos Setoriais." *Inovação Uniemp* 2 (2): 26–27. http://inovacao.scielo.br/scielo.php?script=sci_arttext&pid=S1808-23942006000200014&nrm=iso.

Furtado, André Tosi, and Ruy de Quadros Carvalho. 2005. "Padrões de Intensidade Tecnológica Da Indústria Brasileira: Um Estudo Comparativo Com Os Países Centrais." *São Paulo Em Perspectiva* 19 (1): 70–84. doi:10.1590/S0102-88392005000100006.

Furtado, André Tosi, Ruy Quadros, and Silvia A Domingues. 2007. "Intensidade de P&D Das Empresas Brasileiras." *Inovação Uniemp* 3 (6): 26–27. http://inovacao.scielo.br/scielo.php?script=sci_arttext&pid=S1808-23942007000600013&nrm=iso.

Lauglo, Jon. 2009. “Research for TVET Policy Development.” In *International Handbook of Education for the Changing World of Work SE - 60*, edited by Rupert Maclean and David Wilson, 891–904. Springer Netherlands. doi:10.1007/978-1-4020-5281-1_60.

Neugart, Michael, and Klaus Schömann. 2003. *Forecasting Labour Markets in OECD Countries: Measuring and Tackling Mismatches*. Labour Markets and Employment Policy Series. Edward Elgar.

Nisbet, Robert, John Elder, and Gary Miner. 2009. *Handbook of Statistical Analysis and Data Mining Applications*. 1st ed. Elsevier.

OECD. 2003. *OECD Science, Technology and Industry Scoreboard 2003*. Paris: OECD Publishing. doi:10.1787/sti_scoreboard-2003-en.

———. 2009. “Meeting Labour Market Needs.” *Learning for Jobs*. Paris: OECD. http://www.oecd.org/edu/skills-beyond-school/43898253.pdf.

OpenStax College. 2014. “Demand and Supply at Work in Labor Markets.” *OpenStax-CNX Module: m48637*. http://cnx.org/content/m48637/1.7/.

Rubin, Donald B. 2004. *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley-Interscience.

Silva, Carlos Eduardo Lobo e, and Geoffrey J D Hewings. 2010. “A Decisão Sobre Investimento Em Capital Humano Em Um Arranjo Produtivo Local (APL): Uma Abordagem Teórica.” *Revista Brasileira de Economia* 64 (1): 67–79. doi:10.1590/S0034-71402010000100005.

SPSS. 2001. “The SPSS TwoStep Cluster Component: A Scalable Component Enabling More Efficient Customer Segmentation.” http://www.spss.ch/upload/1122644952_The SPSS TwoStep Cluster Component.pdf.

UNESCO. 2013. “Preliminary Report Accompanied by a First Draft of the Recommendation Concerning Technical and Vocational Education and Training.” Paris. http://unesdoc.unesco.org/images/0022/002296/229649e.pdf.