

TEACHING-LEARNING PROCESS UNDER A MULTIDIRECTIONAL EFFICIENCY AND CANONICAL CORRELATION ANALYSIS

K. Murillo, E. Rocha

University of Aveiro (PORTUGAL)

Abstract

The detailed study of the agents and factors involved in the teaching-learning process is vital to advance in the proposal of strategies that contribute to their improvement in all aspects (methodology, evaluation, resource management, between others). Today, many theoretical studies on the quality of education are known, which makes this a broad topic of discussion. However, education is a universe in constant motion which imposes the necessity of new systematic studies from different and relevant perspectives.

This work aims to contribute to the study of the quality of higher education, evaluating the teaching-learning process, under two approaches. First, examining the technical efficiency of the institutions of higher education (approach 1), and also evaluating the performance of the students of these institutions, in accordance with the influence of the training of the teachers assigned to them (approach 2).

Specifically, the use case study relates to 335 data units representing Colombian higher education institutions, corresponding to 165 institutions in the year 2016 (47 publics and 118 privates), and 170 institutions in the year 2017 (48 publics and 122 privates). The analysis considers variables of training/research for teacher and variables of graduation/performance for students.

For the estimation of the efficiency (approach 1) has used a non-parametric, method based on the Multidirectional Efficiency Analysis (MEA). The study is focused on efficiency score, efficiency ratios and analysis of the output inefficiency index. For the exploration of the relationship, academic qualification/performance (approach 2), is executed a Canonical Correlation Analysis (CCA). Box-Cox transformations and filtering analysis to deal with outliers are applied. The statistical significance of the canonical correlations found, are shown with different statistical tests.

The greatest contribution of this work is the appropriate combination of the MEA and CCA methodologies, which allows a broader vision of what happens in the scheme: institution efficiency/student performance/teacher training. Our results allow us to characterize the institutions in terms of efficiency, discriminating them into seven groups according to the training offer (doctoral programs, masters, specialization, etc.) and comparatively establish the most relevant relationships between teacher training / research and performance of the student.

Keywords: Quality of education, multidirectional efficiency analysis, canonical correlation analysis, teaching-learning process.

1 INTRODUCTION

Perceiving the role that each active agent (as student or teacher), according to its implicit characteristics, interprets in the teaching-learning process, is vital to establish measures that allow improved educational quality in higher education. In this sense, there are many criteria that must be taken into account and that make the difference in the results obtained, such as the pedagogy used by the teacher to teach and the responsibility applied by students to learn. But it is important to keep in mind that there are other variants in this process, such as the existence of a defined study plan according to the knowledge previously acquired by the students (prerequisites) or the physical resources made available by the higher institution to apply different methodologies of teaching. In other words, the importance of the efficiency of the institution in the educational process must also be established, since we cannot ignore that the results of the students could not be the same, if they belong to an institution that does not have a solid structure (for example poor management of resources, overload of physical capacity, etc.) to facilitate the learning process or take it to a higher level.

In this study, we focus our attention under two perspectives: examining the technical efficiency of the higher institution (approach 1) and analysing to what extent teacher training and research influences student results (approach 2), see Figure 1.



Figure 1. Teaching learning process.

In the literature, there are interesting research on the teaching-learning process. We highlight here some results on the approaches outlined above. About approach 1, Hanushek claims that the key to success lies in good teacher training and not in aspects such as in a large investment in education or in the proportion of students, [1]. Instead, for Konsik, research knowledge, knowledge of pedagogies, teaching methods, and processes (reading, writing, and teaching), and knowledge of government initiatives and educational institutions, they are the basis for planning teaching that goes beyond the classroom [1]. In this sense, Thanassoulis et. al, proposed an integrated approach to higher education teaching evaluation that combines the analytical hierarchy process and data analysis, [2]. About approach 2. According to Villa and Villar, established personal interrelationships (between students and teachers within the educational context) are the main tool to build learning, see [3]. For Buendia and Olmedo, the quality of learning is related to the quality of teaching, [4]. Teachers who know their teaching approaches and use them with learning approaches, perform better, according Maquilon [5]. Murillo et. al, used mathematical techniques to show that, there are multivariable correlations between teacher education and student outcomes for a precise set of factors, [6]. For Beck, program planning, professional identity, evaluation and student monitoring are one of the keys to building a quality education [1].

Specifically, this study evaluates the teaching-learning process of 335 Colombian higher education institutions, considering two very different approaches. Approach 1, examines the technical efficiency of the institutions. In this approach the study is considered under different education approaches (doctoral studies, master studies, specialization programs, etc). Approach 2, evaluates the performance of the students of these institutions, in accordance with the influence of the training of the teachers assigned to them. In this approach the study includes indistinctly institutes with different education approaches.

The analysis is made, under an appropriated methodology for each approach. A non-parametric, method based on the Multidirectional Efficiency Analysis (MEA) is used in the approach 1. The study is focused on efficiency score, efficiency ratios and analysis of the output inefficiency index. A Canonical Correlation Analysis (CCA) is executed in the Approach 2. Box-Cox transformations and filtering analysis to deal with outliers are applied. The statistical significance of the canonical correlations found, are shown with different statistical tests. The study considers variables of student results (performance and graduates) and teacher (training and research), defined by the Colombian Model of Performance Indicators of Education (MIDE). For the data set are considered 165 institutions in the year 2016 (47 public and 118 privates), and 170 institutions in the year 2017 (48 public and 122 privates).

This study has an important contribution, using little used exploratory statistical techniques, that allows us to address important questions in improving educational education at a higher level such as: how strong is the correlation that exists between teacher's education and student performance? Are these correlations preserved year by year? To what extent do the correlations between the active agents intervene in the efficiency of education? In what structural way are the three parts related (teacher training, student results and institution efficiency)?

The remainder of the paper is laid out as follows. The next section a brief overview of the methodology used in each approach is given. Section 3 characterizes the data of the study and presents the main results. Finally, in section four, some concluding remarks are formulated.

2 METHODOLOGY

The analysis proposed in this study involves the application of two main techniques: a non-parametric method based on multidirectional efficiency analysis is used to estimate the efficiency (approach 1). A CCA is performed to explore the relationship, academic qualification / performance (approach 2).

2.1 Multidirectional Efficiency Analysis (MEA)

The MEA model also called potential improvement of data envelopment analysis is a non-parametric model appropriate for measuring levels of efficiency, see [7]. The general idea of the MEA model is the following. Let $n=(s, c, t) \in N$ be a tuple identifying the sector $s \in S$, dimension $c \in C$ and year $t \in T$. Considers that any given tuple $n \in N$ produces $J \in N$ outputs $y_j(n)$, $j \in [J]$, using $I \in N$ inputs $x_i(n)$, $i \in [I]$ ($[m]$ denotes the set $\{1, \dots, m\}$, for some $m \in N$). Therefore, $x(n) \in R^I$ is the vector of all the inputs and $y(n) \in R^J$ is the vector of all the outputs. Here the discretionary inputs (their values can be changed) are represented by the first indices d , $1 < d < I$.

For a given data set $Z=\{z(n)\}_N$ with $z(n)=(x(n), y(n))$, the technical efficiency of each institution is measured by calculating the MEA score of each $\bar{n} \in N$, defined as

$$MEA_Z(\bar{n}) = \frac{\frac{1}{\gamma^*(\bar{n})} - \frac{1}{D} \sum_{i=1}^D \frac{x_i(\bar{n}) - \alpha_i^*(\bar{n})}{x_i(\bar{n})}}{\frac{1}{\gamma^*(\bar{n})} + \frac{1}{J} \sum_{j=1}^J \frac{\beta_j^*(\bar{n}) - y_j(\bar{n})}{y_j(\bar{n})}} \quad (1)$$

where $\alpha_i^*(\bar{n})$, $\beta_j^*(\bar{n})$ and $\gamma^*(\bar{n})$ represent the corresponding optimal solutions to the linear optimization problems $P_i^\alpha(z, n)$, $P_j^\beta(z, n)$ and $P^\gamma(z, n, \alpha^*, \beta^*)$:

$P_m^\alpha(z, \bar{n})$:	$P_j^\beta(z, \bar{n})$:	$P^\gamma(\alpha, \beta, z, \bar{n})$:
min $\alpha_m(\bar{n})$ such that	max $\beta_j(\bar{n})$ such that	max $\gamma(\bar{n})$ such that
$\sum_n \lambda_n x_m(n) \leq \alpha_m(\bar{n})$	$\sum_n \lambda_n x_i(n) \leq x_i(\bar{n})$, $i \in [I]$	$\sum_n \lambda_n x_i(n) \leq x_i(\bar{n}) - \gamma(\bar{n})(x_i(\bar{n}) - \alpha_i^*(\bar{n}))$, $i \in [M]$
$\sum_n \lambda_n x_i(n) \leq x_i(\bar{n})$, $i \in [I] \setminus \{m\}$	$\sum_n \lambda_n y_j(n) \geq \beta_j(\bar{n})$, $j \in [J]$	$\sum_n \lambda_n x_i(n) \leq x_i(\bar{n})$, $i \in [I] \setminus \{m\}$
$\sum_n \lambda_n y_l(n) \geq y_l(\bar{n})$, $l \in [J]$	$\sum_n \lambda_n y_l(n) \geq y_l(\bar{n})$, $l \in [J] \setminus \{j\}$	$\sum_n \lambda_n y_l(n) \geq y_l(\bar{n}) + \gamma(\bar{n})(\beta_l^*(\bar{n}) - y_l(\bar{n}))$, $l \in [J]$

with $\lambda \in \Lambda^N = \{\lambda \in R^N: \sum_{n=1}^N \lambda_n = 1\}$.

The problem $P_i^\alpha(z, \bar{n})$ refers to the minimization of inputs, $P_j^\beta(z, \bar{n})$ to the maximization of outputs and $P^\gamma(\alpha^*, \beta^*, z, \bar{n})$ to the realization of both (minimization of inputs and maximization of outputs) simultaneously. The value in equation (1) varies between 0 and 1, with fully efficient institutions having efficiency scores equal to 1. In addition to the calculation of MEA, the equation (2), based on the ideas in [8], is used to know the number of times each output was used inefficiently.

Definition 2 The inefficiency index for each output is given by

$$R_i(\bar{n}) = \frac{\sum_{n=1}^N \gamma(\bar{n}) (\beta_j^*(\bar{n}) - y_j(\bar{n}))}{\sum_{n=1}^N y_j(\bar{n})} \quad (2)$$

for $j \in [J]$ and tuple $\bar{n} \in N$.

2.2 Canonical Correlation Analysis (CCA)

The CCA is a multivariate analysis of correlation used to measure the associations among two sets of variables. The coefficient obtained measures the strength of association between two canonical variates see [9]. The general idea of CCA is the following. Let X be a set of variables with p number of elements and Y a set of variables with q number of elements. Consider U be the set of linear combinations of X , $U_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p$ with $i=1 \dots p$ and V be the set of linear combinations Y , $V_j = b_{j1}Y_1 + b_{j2}Y_2 + \dots + b_{jq}Y_q$ with $j=1 \dots q$. Suppose $p \leq q$ and define (U_i, V_i) as the i^{th} canonical variate pair, where each member of U is paired with a member of V . The canonical correlation for the i^{th} canonical variate pair, between the linear combinations U_i and V_i , is given by

$$\rho_i = \frac{\text{cov}(U_i, V_i)}{\sqrt{\text{var}(U_i)\text{var}(V_i)}} = \frac{\sum_{k=1}^p \sum_{l=1}^p a_{ik} b_{il} \text{cov}(X_k, Y_l)}{\sqrt{(\sum_{k=1}^p \sum_{l=1}^p a_{ik} a_{il} \text{cov}(X_k, X_l))(\sum_{k=1}^p \sum_{l=1}^p b_{ik} b_{il} \text{cov}(Y_k, Y_l))}}, \quad (3)$$

where $\text{cov}(U_i, V_i)$ denotes the covariance between U_i and V_i and $\text{var}(U_i)$ denotes the variance of the variable U_i . (similar to V_i). The CCA consists in to find linear combinations of the X 's and linear combinations of the Y 's that maximize the correlation in (3). This mean, to find the coefficients of U_i and V_i that maximize the canonical correlation ρ_i subject to the constraints that

$$\begin{aligned} var(U_i) = var(V_i) = 1; cov(U_1, U_i) = cov(V_1, V_i) = 0; cov(U_2, U_i) = cov(V_2, V_i) = 0, \dots \\ cov(U_i - 1, U_i) = cov(V_i - 1, V_i) = 0, cov(U_1, V_i) = cov(U_i, V_1) = 0, cov(U_2, V_i) = cov(U_i, V_2) = 0, \dots \\ cov(U_i - 1, V_i) = cov(U_i, V_i - 1) = 0. \end{aligned}$$

with all the remaining correlations equal zero.

In this approach, are also applied techniques as the use of Box-Cox transformations, to modify the distributional shape of a set of data, see [10]. In order, to test the i th canonical correlation and all that follow it are zero, is used statistical tests based on Wilks Lambda, Hotelling-Lawley Trace, Pillai-Bartlett Trace and Roy's Largest Root, see [11].

3 DATA CHARACTERIZATION AND DISCUSSION RESULTS

3.1 Data characterization

The proposed analysis is applied to 335 Colombian HEIs (higher education institutions), 165 (2016) and 170 (2017), which are divided into seven academic sectors (see Table 1). For the purposes of this study, indicators of the quality of public and private HEIs are considered, determined by the Model of Educational Performance Indicators (MIDE), see [12]. The Colombian model includes three main dimensions (students, teaching, environment) associated with educational quality. We studied the variables considered by the multidimensional MIDE model in 2016 and 2017, following the structure presented for each year (see Table 2). The data set includes information supplied by the Colombian ministry.

Table 1. Sectors description.

	<i>Label</i>	<i>Description</i>	<i>2016</i>	<i>2017</i>
S1	Doctoral	Institutions with a doctorate focus	7	8
S2	Master	Institutions with a focus on master's degrees	31	14
S3	Univ. 5-8	Universities with 5 to 8 areas of knowledge.	16	31
S4	Other 5-8	Institutions with 5 to 8 areas of knowledge that are not in S3.	49	29
S5	Univ. 2-4	Universities with 2 to 4 areas of knowledge.	18	13
S6	Other 2-4	Institutions with 2 to 4 areas of knowledge that are not in S5.	14	58
S7	Specialized	Institutions with undergraduate degrees in a specialized area of knowledge.	30	17

Table 2: Variables description

	<i>Teacher education/research outcomes</i>	<i>Dimension</i>	<i>Students results</i>
2016	V01 Undergraduate students*	Performance	V08 Quantitative reasoning
	V02 Teachers with masters		V09 Critical reading
	V03 Teachers with doctorate		V10 Written communication
	V04 Teachers with specialization	Graduates	V11 Graduates-undergraduate
	V05 Teachers without postgraduate		V12 Employability
	V06 Citations		V13 Step to postgraduate
	V07 Intellectual production score		
2017	V14 Undergraduate students*	Performance	V20 Quantitative reasoning
	V15 Teachers with masters		V21 Critical reading
	V16 Teachers with doctorate		V22 Written communication
	V17 Teachers without postgraduate	Graduates	V23 Graduates OLE Employability
	V18 Citations		V24 Employability
	V19 Article		V25 Step to Postgraduate

* Variable used to normalize the other data.

3.2 Approach 1 based on MEA

For an analysis more detailed, in this approach we consider the student results divided into two dimensions: Performance and Graduates (see Table 2). We calculated the MEA score for each institution in each year, considering the variables of teacher education and research outcomes as the inputs and the students results as the outputs.

Defined EFF as subset of sector/dimension/year ternaries $n=(s, c, t)$ such as $0,6 \leq \text{MEA}(s, c, t) \leq 1,0$. The total mean efficiency by each sector is provided in the Table 3, and the percentage of institutions with MEA efficiency score equal to 1 (Full-EFF) are presented in Figure 2.

Table 3. Total mean efficiency by sector (2016-2017).

		S1	S2	S3	S4	S5	S6	S7
2016	Performance	0.4	0.2	0.4	0.2	0.6	0.4	0.1
	Graduates	0.6	0.7	0.8	0.6	0.8	0.8	0.6
2017	Performance	0.4	0.4	0.1	0.1	0.3	0.1	0.4
	Graduates	0.8	0.9	0.6	0.5	0.9	0.4	0.7

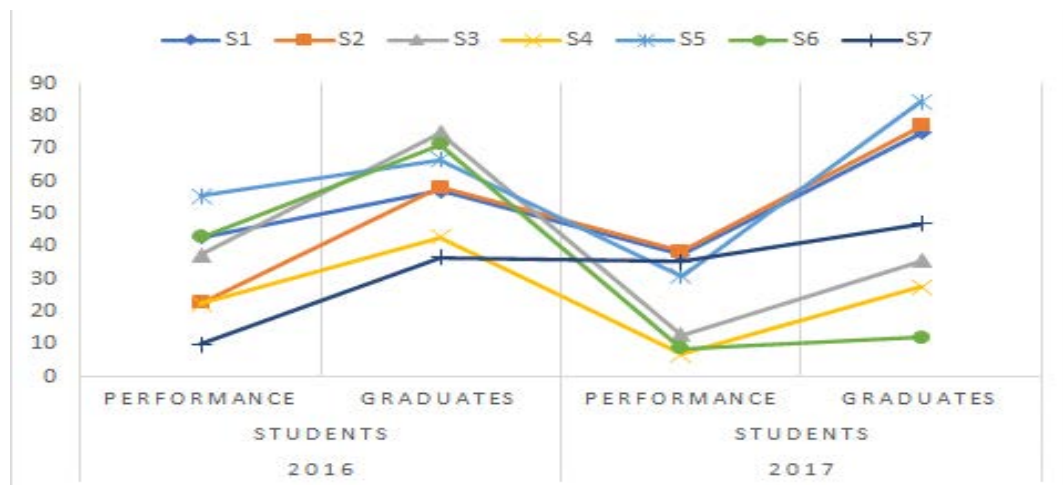


Figure 2. Full-EFF by sector (2016-2017).

Table 3 and Figure 2, allow us to see the behaviour of the sectors in each dimension, throughout the study period. The performance dimension is less efficient than graduates in both years. S5 is a sector with good performance. In 2016 are the sectors with the highest average total efficiency in graduates (2017), performance and graduates (2016).

The values (percentages) in Figure 3 represent the number of times, in which each output was used inefficiently in each sector (see equation 2). In the Figure is represented the output (first, second and third) for each dimension according to the order established in the Table 2. Paradoxically, although some variables were modified (see Table 2) and the number / distribution of institutions is not the same (see Table 1), we found many similarities in the results of the two years. The three outputs in the dimension performance were equally misused and in the graduated dimension, the order of inefficiency output is the same: second, third, and first output.

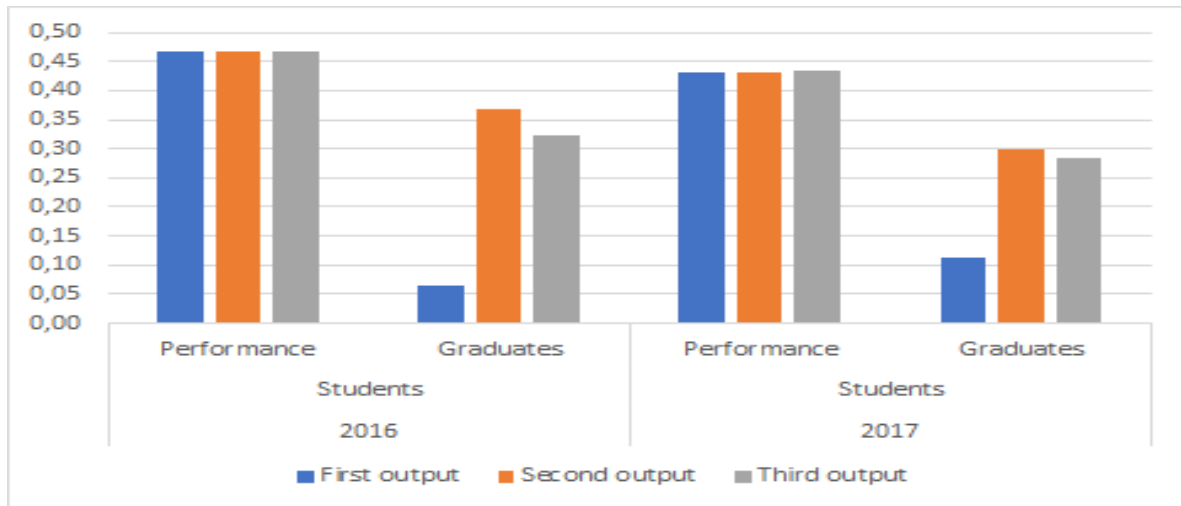


Figure 3. Inefficiency index, sector S1.

3.3 Approach 2: CCA

We used a CCA to establish the correlation between the academic qualification of teachers of HEI (groups of variables Y) and results of their students (group of variables X) in Colombian.

We normalized the variables dividing by the number of students and then apply a filter to shoot the outliers. In the Figure 4, the normality and correlations of the data in the year 2016, before and after of applied of these two techniques are showed. In this case, the filter used was $V04 > 0.9$ and $V06 > 1.0$, reducing the sample in this year to 162 institutions

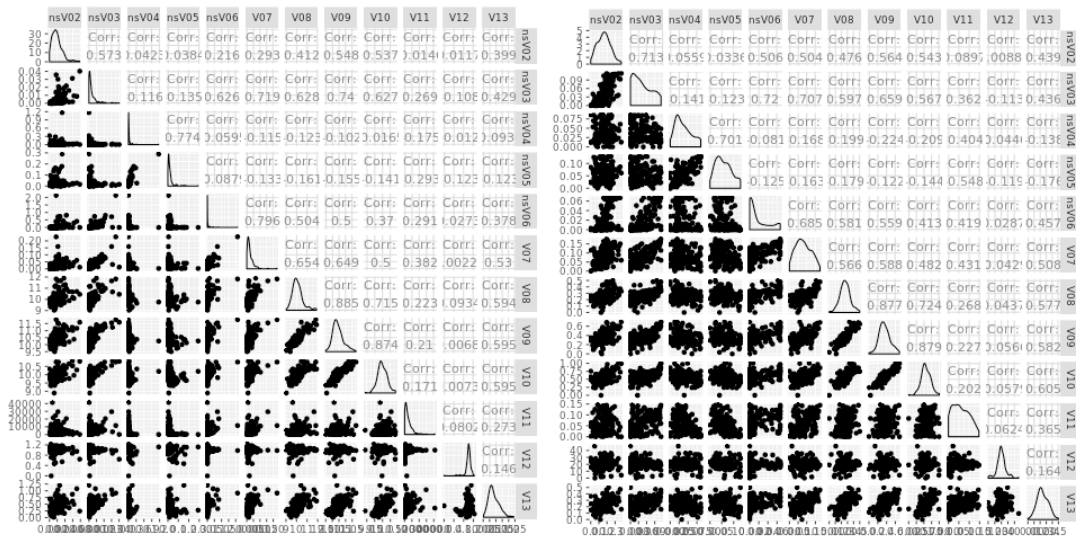


Figure 4. Normality and correlations, 2016: initial data (left side); preprocessed data (right side).

Since the multivariate normality of the variables is required, we need apply the Box-Cox transformation. For this, we need guarantee that the values are positive, and it is possible using the calculation in equation (4).

$$V_{ix}^* = \frac{V_i - \min(V_i)}{\max(V_i) - \min(V_i)} \quad \text{and} \quad V_{jy}^* = \frac{V_j - \min(V_j)}{\max(V_j) - \min(V_j)} \quad (4)$$

for $i=02, \dots, 07$ and $j=08, \dots, 13$ (year 2016) and for $i=15, \dots, 19$ and $j=20, \dots, 25$ (year 2017).

For the interesting reader present to continue, the correspondent power parameters lambda in the transformations in 2016: 2.000, -8.810, -11.639, -7.719, -14.913 and -5.551 for V_{ix}^* with $i= 02, \dots, 07$. For V_{iy}^* with $i=08, \dots, 13$ the values are -1.000, 0.000, 1.000, -6.752, 8.651 and -1.000. In the next year

(2017), the values of V_{ix}^* with $i=15, \dots, 19$ are -5.220, -2.000, -7.424, -13.200 and -14.521. For V_{iy}^* with $i=20, \dots, 25$ the correspondent values are 0.000, 0.000, 1.000, -5.462, 2.000 and -3.739 respectively.

We apply the transformation in all the variables, and we standardized the variable V_x^* (and V_y^*) by using \widehat{V}_{ix}^* (resp. \widehat{V}_{iy}^*), defined by:

$$\widehat{V}_{ix}^* = \frac{V_{ix}^* - \overline{V_{ix}^*}}{\text{sd}(V_{ix}^*)} \quad \text{and} \quad \widehat{V}_{iy}^* = \frac{V_{iy}^* - \overline{V_{iy}^*}}{\text{sd}(V_{iy}^*)} \quad (5)$$

where $\overline{V_{ix}^*}$ and $\overline{V_{iy}^*}$ represent the average and $\text{sd}(V_{ix}^*)$ and $\text{sd}(V_{iy}^*)$ the standard deviation (the same form is used for V_{iy}^*).

The heat correlation graphs are shown in Figure 5 and Figure 6. Here the colour intensity is proportional to the Pearson correlation values. The existing correlations between the variables can be perceived using the correct interpretation of the multicolour code: negative correlation (represented by the blue colour) to the positive correlation (red) and null correlation (green). In this way we find similarities and differences in the correlations from year to year (2016: X correlation (between 6 variables), Y correlation (between 6 variables) and 2017: X correlation (between 5 variables), Y correlation (between 6 variables), see Table 2). The use of colour allows us to easily perceive the great differences between the initial data and the processed data in each year and at the same time from year to year. Note that the graphs reflect a higher negative correlation in the processed data.

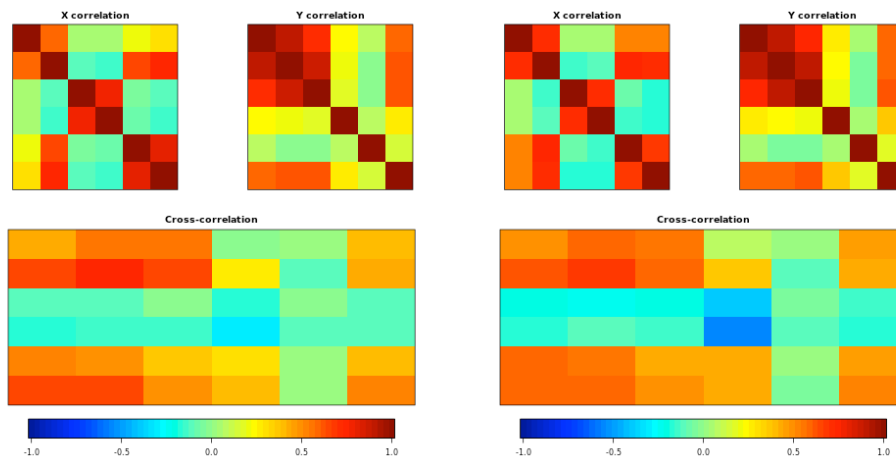


Figure 5. Generating correlation 2016: initial data (left side); preprocessed data (right side).

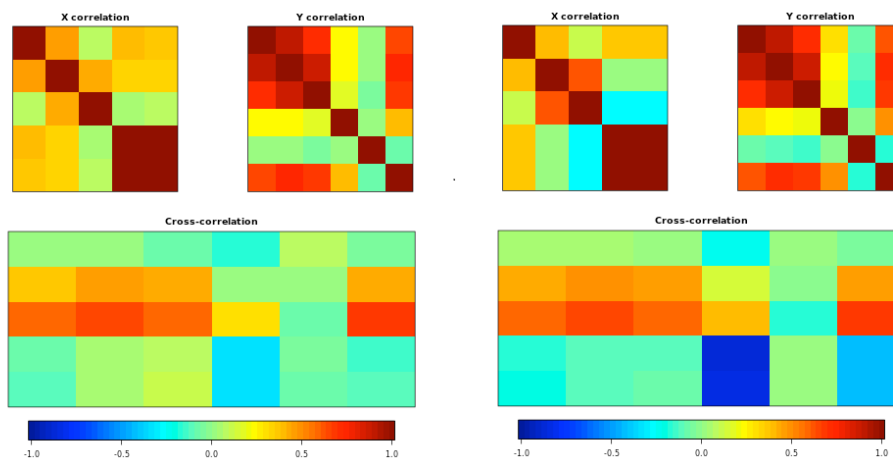


Figure 6. Generating correlation 2017: initial data (left side); pre-processed data (right side).

The canonical variate pairs were analysed for each year and the results of the five statistical tests applied are provided in Table 4. As we can see, in both years, considering 5% as the significance level, only the two last canonical variate pairs are not significantly correlated.

Table 4. Statistics tests (2016-2017).

Wilks' Lambda, using F-approximation (Rao's F)										
	2016					2017				
$H_0^{(i)}$	<i>stat</i>	<i>approx.</i>	<i>df1</i>	<i>df2</i>	<i>p.value</i>	<i>stat</i>	<i>approx.</i>	<i>df1</i>	<i>df2</i>	<i>p. value</i>
i=1	0.193	8.389	36	665.848	0.000	0.028	29.948	30	626.000	0.000
i=2	0.537	4.117	25	566.156	0.000	0.141	20.942	20	521.660	0.000
i=3	0.779	2.481	16	468.060	0.001	0.547	8.907	12	418.320	0.000
i=4	0.872	2.395	9	374.946	0.011	0.973	0.733	6	318.000	0.622
i=5	0.963	1.465	4	310.000	0.212	0.997	0.207	2	160.000	0.813
i=6	0.989	1.688	1	156.000	0.195					

Hotelling-Lawley Trace, using F-approximation										
	2016					2017				
i=1	2.489	10.327	36	896	0.000	7.652	39.380	30	772	0.000
i=2	0.711	4.305	25	908	0.000	3.672	28.718	20	782	0.000
i=3	0.260	2.498	16	920	0.000	0.804	10.614	12	792	0.000
i=4	0.141	2.442	9	932	0.009	0.028	0.743	6	802	0.615
i=5	0.037	1.489	4	944	0.203	0.002	0.209	2	812	0.811
i=6	0.010	1.724	1	956	0.189					

Pillai-Bartlett Trace, using F-approximation										
	2016					2017				
i=1	1.188	6.419	36	936	0.000	2.005	17.850	30	800	0.000
i=2	0.548	3.812	25	948	0.000	1.206	12.869	20	810	0.000
i=3	0.237	2.472	16	960	0.001	0.464	6.993	12	820	0.000
i=4	0.130	2.409	9	972	0.010	0.027	0.756	6	830	0.604
i=5	0.037	1.528	4	984	0.191	0.002	0.216	2	840	0.805
i=6	0.010	1.780	1	996	0.182	2.005	17.850	30	800	0.000

Roy's Largest Root, using F-approximation:										
	2016					2017				
1	0.640	46.237	6	156	0	0.799	106.115	6	160	0

Now it is necessary testing the significance of each canonical correlation (X_c), see Table 5. The estimates of the first canonical correlations, associated to variable pairs with statistically significant correlations are: $r_1=0.800$, $r_2=0.557$, $r_3=0.326$, and $r_4=0.306$ in the 2016; and $r_1=0.894$, $r_2=0.861$ and $r_3=0.661$ in the 2017. Considering the correlation square root (R^2), that's mean, that canonical variate pairs with significant values for the determination coefficient are only the first two and third correlations in 2016 and 2017, respectively. In fact, the 64.0% 31.0% 10.6% of the variation of U_1 , U_2 are explained by V_1 , V_2 and V_3 in 2016. The 79.9%, 74.1% and 43.7% of the variation of U_1 , U_2 and U_3 are explained by V_1 , V_2 and V_3 in 2017, respectively.

Table 5. Test for significance of each canonical correlation (2016-2017).

	2016					2017				
	<i>CanR</i>	<i>LR test stat</i>	<i>Approx F</i>	<i>Num DF</i>	<i>Pr(>F)</i>	<i>CanR</i>	<i>LR test stat</i>	<i>Approx F</i>	<i>Num DF</i>	<i>Pr(>F)</i>
1	0.800	0.193	8.389	36	<2.2e16	0.894	0.028	29.948	30	<2.2e-16
2	0.557	0.537	4.117	25	0.000	0.861	0.141	20.942	20	<2.2e-16
3	0.326	0.779	2.481	16	0.001	0.661	0.548	8.907	12	0.000
4	0.306	0.872	2.395	9	0.011	0.157	0.973	0.733	6	0.623
5	0.162	0.963	1.465	4	0.212	0.051	0.997		2	
6	0.103	0.989	1.688	1	0.195					

To finish the analysis, we need to carry out two important steps. Identify the contributions of the individual variable to each corresponding canonical variable X_c and Y_c (i.e. the magnitudes of the coefficients) and estimate the correlation coefficients between canonical variables and initial variables. The results of the two steps for 2017, can be seen in Tables 6 and Table 7.

Table 6. Standardized coefficients, 2017.

	X_{c1}	X_{c2}	X_{c3}	X_{c4}	X_{c5}		Y_{c1}	Y_{c2}	Y_{c3}	Y_{c4}	Y_{c5}
V15	0.06	-0.04	-0.19	1.34	-0.45	V20	-0.94	-1.88	-0.33	0.34	-0.03
V16	0.22	-0.11	-0.89	-0.99	0.16	V21	0.79	1.99	-0.41	-0.38	-0.45
V17	-0.03	-0.09	0.14	-0.07	1.16	V22	0.01	-0.15	-0.03	0.25	-0.80
V18	-17.12	-24.2	-3.61	2.81	-4.06	V23	0.80	-0.28	0.52	-0.14	-0.59
V19	16.41	24.75	3.04	-3.04	3.77	V24	0.09	0.13	0.14	0.99	0.11
						V25	0.26	-0.21	-0.37	0.18	1.50

Table 7. Correlations, 2017.

Correlations of X with X_c						Correlations of Y with Y_c					
	$xc1$	X_{c2}	X_{c3}	X_{c4}	X_{c5}		Y_{c1}	Y_{c2}	Y_{c3}	Y_{c4}	Y_{c5}
V15	0.25	-0.02	-0.67	0.68	0.13	V20	0.15	-0.46	-0.79	0.18	-0.27
V16	0.48	-0.24	-0.83	-0.11	0.09	V21	0.36	-0.02	-0.87	0.12	-0.26
V17	-0.23	0.08	-0.22	0.28	0.90	V22	0.37	0.04	-0.78	0.12	-0.34
V18	-0.78	0.55	-0.28	0.03	0.05	V23	0.85	-0.47	0.14	-0.01	-0.15
V19	-0.77	0.58	-0.28	0.03	0.06	V24	-0.00	0.11	0.27	0.95	-0.01
						V25	0.64	-0.18	-0.66	0.03	0.34

Higher correlation values (in absolute value) for the first canonical variables in X are found. In 2016 with V03, V06 and V07 in the range [0.69,0.83]. For the second canonical variable, with V02 and V05, [0.56, 0.68]. In 2017 between V18 and V19, for the first and the second canonical variable in [0.65,0.78] and [0.48, 0.57] respectively. The correlation values for the first canonical variable in Y are stronger in 2016 with V8, V9, V10 and V11, [0.67,0.81]; and in 2017 with V20, V21 and V22, [0.67,0.81].

4 CONCLUSIONS

The present work uses the MEA and CCA techniques, with the aim of contributing to the improvement of educational quality at the higher level, under two approaches: examining the technical efficiency of institutions (approach 1), and establish the relationships between student performance and teacher training (approach 2). The analyses applied to Colombian higher education institutions between the years 2016-2017, showed interesting results in each approach. Universities more specialized with 2 to 4 areas of knowledge, obtained higher levels of efficiency in relation to other universities. According to the teacher training, the results in quantitative reasoning critical reading, and written communication can be improved by almost 60%. In general terms, the "performance" dimension performs better throughout the study period than the "graduates" dimension. Regarding approach 2, the results are maintained year by year (2016-2017), with little variation in the two types of correlations found: direct and by groups. Among the most outstanding we find direct correlations (in the same group of variables) between critical reading and written communication; quantitative reasoning and critical reading; and teachers with a doctorate and the number of citations. Correlations by groups of factors: teachers with doctoral training and high experience in research on students' cognitive abilities (quantum reasoning, communicative reading, and written communication); teachers with master's degrees on students' ability to read critically, internal scores established by the MIDE model (V11, V23) and written communication.

All the above shows the complexity of the teaching-learning process and the importance of using competent tools to adequately examine and evaluate each of the different factors involved. Let us not forget that in educational institutions it is important to keep up with the needs and resources available

today, to provide strategies that allow students to develop their critical and analytical capacity, preparing them to be an active part of an increasingly competent society.

ACKNOWLEDGEMENTS

Work partially supported by CIDMA through the Portuguese Foundation for Science and Technology (FCT - Fundação para a Ciência e a Tecnologia), references UIDB/04106/2020 and UIDP/04106/2020. Murillo is also supported by national funds (OE), through FCT, I.P., in the scope of the framework contract foreseen in the numbers 4, 5 and 6 of the article 23, of the Decree-Law 57/2016, of August 29, changed by Law 57/2017, of July 19.

REFERENCES

- [1] Communication from the Foro Maestro Siempre, “La calidad en la educación depende directamente de la calidad de los profesores, los educadores y los currículos”, May 5, Bogotá, 2014, available at <https://www.mineducacion.gov.co/cvn/1665/w3-article-340967.html>.
- [2] E. Thanassoulis, P. K. Dey, K. Petridis, I. Goniadis and A.C. Georgiou, “Evaluating higher education teaching performance using combined analytic hierarchy process and data envelopment analysis”, *Journal of the Operational Research Society*, vol. 68, pp. 431–445, 2017.
- [3] A. Villa and L. M. Villar, “Clima organizativo y del aula”, Servicio Central de Publicaciones. Vitoria, Gasteiz: Gobierno Vasco. pp. 333-353, 1992.
- [4] L. Buendía and E.M. Olmedo, “Estudio transcultural de los enfoques de aprendizaje en Educación Superior”, *Revista de Investigación Educativa*, vol. 21, no. 2, pp. 371-386, 2003.
- [5] J.J. Maquilón, “Diseño y evaluación del diseño de un programa de intervención para la mejora de las habilidades de aprendizaje de los estudiantes universitarios”, Murcia: EDITUM, 2010.
- [6] K. Murillo, E. Rocha, A. Freitas, C. Pardo, “Does teacher's education affect student performance?” *EDULEARN19 Proceedings*, pp. 5899-5908, 2019.
- [7] P. Bogetoft, and J.L. Hougard, “Efficiency evaluations based on potential (non -proportional) improvements, *J. Productivity Analysis*, vol. 12, no. 3, pp. 233-247, 1999.
- [8] P. Bogetoft, and L. Otto, “Benchmarking with DEA, SFA and R”, *International Series in Operations Research and Management Science*, vol. 157, pp. 1–368, 2011.
- [9] H. Hotelling, "Relations between two sets of variates", *Biometrika*, vol. 28, no. 3–4, pp. 321–377, 1936, doi:10.1093/biomet/28.3-4.321. JSTOR 2333955.
- [10] G. Chen, R. A. Lockhart and M. A. Stephens, “Box–Cox transformations in linear models: large sample theory and tests of normality”, *The Canadian Journal of Statistics* 1, vol. 30, no. 2, pp. 1–59, 2002.
- [11] D.L. Hahs-Vaughn, “Applied multivariate statistical concepts”, 1st Edition, New York, pp. 1-662, 2016, <https://doi.org/10.4324/9781315816685>.
- [12] Comunicación del Gobierno de Colombia, MIDE: Modelo de Indicadores del Desempeño de la Educación, Documento metodológico MIDE Universitario, 1-52, 2017.