

Factors Driving University Choice: A Principal Component Analysis on Italian Institutions

Azzone, Giovanni; Soncin, Mara

Department of Management, Economics and Industrial Engineering, Politecnico di Milano, Italy.

Abstract

When investigating students' motivations to enroll in university, a wide range of elements related to the overall student experience concerning both the institution and the surrounding context should be taken into account. The current study moves from this point to analyse students' choice factors from a survey completed by 27,705 students across 23 Italian institutions by means of a logistic Principal Component Analysis. Results confirm the presence of multiple factors jointly influencing students' choice, with geographical proximity, job opportunities in the region, university reputation and ease of access opposing one another. Aggregating results at institutional level, students' distribution prove to be highly heterogeneous across universities. From this, a managerial tool is provided to position student population and derive strategic implications. Finally, policy considerations are reported.

Keywords: Higher education; choice factors; principal component analysis.

1. Introduction and background

Higher Education (HE) institutions' capacity to create public value is intrinsically embedded in their interaction with students, who are not only the final recipients of service delivery, but also play an active role in value production by means of an iterative interplay that place them at the center of the service *experience* (Osborne *et al.*, 2015). Under the assumption that value increases when users are involved in its creation, a clear understanding of their expectations facilitate the process and maximise potential results, enhancing the role of public HE institutions as *public value co-creators* (Osborne *et al.*, 2016). To enforce the co-creation process, a clear understanding of students' motivations is a necessary condition. First year students represent a case in point in this respect, given the tense expectations they have when they enter tertiary education (Trotter & Roberts, 2006). These are not only expectations in terms of teaching and learning quality and support (Jamelske, 2009), but recall a *total student experience* approach (Petruzzellis & Romanazzi, 2010). Indeed, the current study focuses on the student as unit of analysis, trying to identify the dimensions of the university experience that are mostly taken into account when choosing the HE institution. Overall, a better understanding of the drivers of choice in a total student experience approach makes a step towards the co-creation of public value, aligning students and HE institutions' perspectives. According to the literature on the topic, factors influencing students' university choice have been evolving over time. A report by the British Council (2017) affirms that in the UK students now consider availability of scholarships as the most important factor, while university ranking has dropped to the last position. In their study, Simões & Soares (2010) find that geographical proximity is the factor mainly considered by students in a Portuguese institution, followed by the university reputation. Briggs and Wilson (2007) examine the effect of costs and information in six Scottish universities, finding that course content information play a prominent role in students' choice. However, most of the cited studies do not completely exploit the complexity of students' decision that may jointly consider multiple factors related to different aspects of their university experience. In additional to that, the variability in the distribution of students' choice across institutions has high potential implications that are currently under-investigated. This study moves from these aspects analysing the case of Italy, where the HE system is characterised by a particularly challenging context in terms of number of young people (25-34 years old) earning a degree, which is 24% compared to an OECD average of 41%, or in terms of bachelor's students dropping out (13.9%) or switching to a different major or university (15.4%) by the end of the first year (ANVUR, 2016). Understanding the process that brings students to the university choice has major implications at managerial and policy level. For HE institutions' managers, understanding their students' motivations may bring to a greater strategical awareness of their attractiveness in order to better shape their orienteering activities towards either the current

or a more “desirable” population of students. At policy level, it may help to reduce the proportion of students dropping out or switching from the university where they originally enrolled by means of evidence based policy addressed to different student populations. Hence, the research questions can be stated as:

- *Which are the main multidimensional factors that drive students’ university choice in the Italian context?*
- *Is there a different distribution of these factors across HE institutions?*

For this purpose, a logistic Principal Component Analysis (PCA) has been applied to data on students’ choice collected by surveying 27,705 students across 23 public HE Italian institutions. The originality of the research lies in the creation of a unique dataset of institutions representative at national level, which allows to make considerations about students, but also to aggregate students’ choice at university level. As a remainder of the paper, section 2 reports data and methodology, while sections 3 and 4 respectively present and discuss the results.

2. Data and methods

2.1. Data

Data used in the analysis are collected within a larger project self-financed by Italian public universities called *Good Practice project*. In the edition 2016/17, 38 HE institutions took part in the project that aims at benchmarking the performances of institutions in terms of efficiency (cost per unit of output) and perceived effectiveness (surveys filled out by the main stakeholders). In the wave here considered, 23 HE institutions agreed to join the student survey. The first-year students survey is enriched by a set of additional questions concerning the factors driving the choice of the university to attend. These are related to a cross definition of university performance, which combines the characteristics of the HE institutions with those of the surrounding social and economic fabric. In details, options are related to:

- *economic factors* such as tuition fees, scholarships and financial aid provided by the university;
- *reputational factors*, related to the university prestige both in terms of “word-of-mouth” and official rankings;
- *ease of access*, linked to the presence and difficulty level of the entrance test and to the prerequisites needed to successfully attend the programme;
- *quality of student services* such as availability of information, quality of the facilities and orienteering activities;
- *proximity* to the home town;

- *quality of life* in terms of public services, amenities and metropolitan area;
- *job opportunities* in the region where the university is located, in terms of average wage, employment level and proximity to industrial areas;
- *cost of life* in the area where the university is located.

The survey was anonymous and administered over a period of nearly one month at the end of the first year of students' attendance. It is not mandatory for students to fill the questionnaire out, so the number of respondents varies across institutions. It is worth to notice that the analysis has been run at student level and later on re-aggregated at institutional level to investigate how students types are distributed across universities, so that the principal components definition is not influenced by the representativeness at institutional level. As a descriptive, Table 1 reports the number of respondents by institution and the response rates (calculated as the number of respondents over the total number of first-year students).

Table 1. List and response rate of participant universities.

University	N	Response rate	University	N	Response rate
A	2,482	13%	M	349	9%
B	1,533	54%	N	1,028	21%
C	2,448	55%	O	1,395	8%
D	336	4%	P	191	5%
E	1,814	64%	Q	157	5%
F	245	33%	R	417	15%
G	1,435	19%	S	1,183	9%
H	1,191	11%	T	3,287	86%
I	980	16%	U	407	11%
J	340	7%	W	694	16%
K	655	19%	X	5,492	70%
L	20	14%	Total	27,705	

In the survey, the student is not forced to choose only one driver of choice, but can select from one to three options, under the assumption that student's choice is multidimensional in nature and, hence, a typology of students' choice can be analysed. Table 2 reports summary statistics about the variables considered. Proximity to home is the factor selected by the highest percentage of students, showing the high costs (monetary and non-monetary) of mobility even within the country. University reputation is the second most selected factor, highlighting the importance of university prestige. Hence, the descriptive analysis confirms the results from the literature, as reported by Simões & Soares (2010).

2.1. Methodology

The methodology applied for data analysis is *logistic Principal Component Analysis (PCA)*. Logistic PCA is a method for dimensionality reduction of binary data, which moves from the original formulation of PCA by Pearson (1901) to take into consideration the possible dichotomous nature of data (Landgraf & Lee, 2015). The model is based on the definition of the best projection of parameters from a saturated model (namely a model where the

number of parameters matches the number of data points), in order to minimise the deviance (that is to minimise the distance from the overfitting saturated model). As in traditional PCA, the selection of the number of principal components is based on different parameters, namely the cumulative variance explained by the components, the existence of an elbow in the scree plot representing the variance explained by the components (for example, the presence of an elbow between $k-1$ and k components may lead to consider $k-1$ principal components) and the distribution of the scores across components, again looking for a change in the distribution when increasing the number of components. Once that the best number of dimensions has been set, *component loadings* allow for their interpretation, providing a measure of the correlation between the original variables and the new components. In order to facilitate their interpretation, a *varimax rotation* has been applied to component loadings, orthogonally rotating the reference system to minimise the number of factor loadings significantly contributing.

Table 2. Descriptive statistics of the factors driving university choice.

Variable	N	Percentage of students selecting the variable
Economic factors	4,289	15%
Reputation	9,551	34%
Ease of access	5,388	19%
Quality of student services	3,260	12%
Proximity to home	12,268	44%
Quality of life	2,888	10%
Job opportunities in the area	6,275	23%
Cost of life	1,293	5%

Note: Sum of percentages is greater than 100% given the possibility to select more than one variables per respondent.

Using the methodology here described, we are able not only to reduce the variables concerning university selection to a smaller number of dimensions better explaining the phenomenon observed, but also to demonstrate the interrelation existing between the different dimensions of choice, making latent patterns emerge from data. The existence of more than one component is able to demonstrate that students do consider a varied and interrelated spectrum of factors when they choose the university to attend.

3. Results

The number of components from the application of logistic PCA to the 27,705 student observations has been selected considering multiple indicators. The box plot representing the distribution of the scores along components reports an elbow between the third and the fourth component, suggesting that three is most suitable number of factors to be considered. With respect to the saturated model, the first component counts for the 36.9% of the total

variance. The second and third components respectively represents the 16.4% and 13.6% of the variance for a total amount of 66.9% of variability explained, which suggest a satisfying fit of the component solution. The loadings for the three principal components are represented in Figure 1. The *varimax* rotation reduces to zero some of the factor loadings to facilitate the interpretation of those that contribute the most to the component interpretation. The first component, which is explaining more than one third of the total variance, is actually a combination of the different variables. In this sense, this confirms that students do have a multifaceted approach to university choice. The fact that, among variables, the cost of life in the area report the highest absolute value highlights the important role played by the surrounding conditions when choosing university. The second component is instead represented by the contraposition of two variables, which have the highest loading value: proximity to home (PR) on one side and job opportunities in the area (JO) on the other. Hence, students are more inclined to move to a different region if they consider that attractive in terms of job opportunities. This a very *long-term perspective*, which highlights the wide vision students may adopt when choosing university. On the other hand, students that choose an institution for its proximity to the home town are not particularly interested in the implications in terms of employability in the long run. This a more *now and here* vision, which may be related to a number of physical, financial and social constraints. Finally, the third principal component is mainly explained by two variables, again pointing to opposite directions: reputation of the institution (RE) and ease of access (EA). From this point of view, students who choose an institution because of its national and international prestige do not care about how difficult it is to be enrolled. Their motivation is much more *emotional* and related to the social dimension of being part of a prestigious organisation. By contrary, part of students consider the low entrance barriers as an important driver of choice. They may consider this as a way to *minimise failure possibilities* and hence to increase their ability to succeed in studies later on. Aggregating the component scores at institutional level, it is possible to represent the university population according to the main drivers of choice. In terms of managerial implications, this may be used by institutions to have a deeper understanding of their students' population in order to reinforce their attractiveness in this direction or to modify the target population. In both cases, awareness of students' perception is a fundamental point. As examples of possible maps to be created at institutional level, Figure 2 reports students distribution along components that have a remarkable contrast between dimensions: the second (x axis) and the third components (y axis). Examples for three institutions are here represented to stress differences across universities that are located in the same region, with similar contextual characteristics. This provides evidence about the fact that even within the same context, factors of choice may differ.

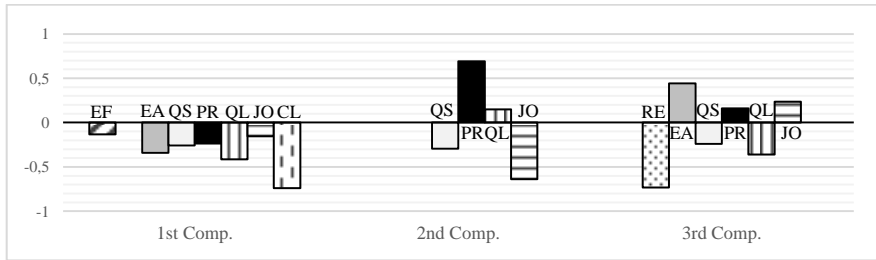


Figure 1. Factor loadings along the three principal components.

Note: EF=Economic Factors; RE=Reputation; EA=Ease of Access; QS=Quality of Students services; PR=Proximity; QL=Quality of Life; JO=Job Opportunities; CL=Cost of Life.

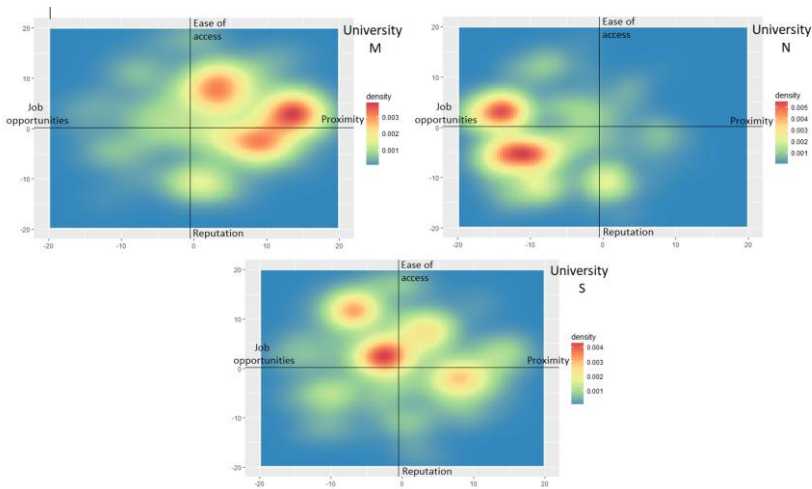


Figure 2. Students distribution along component two (Job opportunities vs. Proximity) and component three (Reputation vs. Ease of access).

Note: A higher density (red colour) represents a higher number of students with similar scores on that component.

4. Discussion

The current study moves from the idea that students may consider a composite range of factors that are widely related to the student experience when choosing university, and that a clear understanding of these dimensions align students and HE institutions’ perspective facilitating the co-creation of public value. Applying a logistic PCA to a dataset of 27,705 students reporting the factors driving their university choice in 23 Italian HE institutions, composite elements emerge. Moving from the first component where the different factors are jointly considered in student’s choice, four main dimensions are opposed: proximity to the home town *versus* job opportunities in the region on the one hand; university reputation *versus* ease of access on the other hand. Looking at the graphical maps (Figure 2) a great

variation in students' distribution across HE institutions emerge. For instance, a high concentration of students selects University M because of its proximity to home and ease of access (red areas, with a higher density). However, some students also consider its reputational impact as the main driver (yellow area in the bottom). Hence, a first consideration relates to the possibility for the HE institutions to implement strategic actions to leverage on the current drivers to know better their student population, attracting students aligned with their strategic vision or moving towards different and more "desirable" factors of choice. As a second point, it is worth to notice that students' distribution is quite sparse along the four dimensions. Hence, the population of students is not only differentiated among but also within institutions. This may bring a higher "internal" personalisation of services offered to students with different attitudes and motivations. At policy level, this kind of information may be used to cluster similar institutions across the country, moving towards a higher personalisation of university policies where the institution and the students are at the center of value creation. In this respect, a limitation and a future development of the study is related to the current impossibility to follow students over time and analyse the relationship between university choice and subsequent academic path or drop out probability. Moreover, students' personal information (such as previous career or socio-economic status) is not collected to guarantee complete anonymity, with limitations on the possibility to better characterise student population and to understand the possible link between motivations and social factors.

References

- ANVUR. (2016). Rapporto sullo stato del sistema universitario e della ricerca 2016.
- Briggs, S. & Wilson, A. (2007) Which university? A study of the influence of cost and information factors on Scottish undergraduate choice, *Journal of Higher Education Policy and Management*, 29:1, 57-72.
- British Council. (2017). Ten trends - Transformative changes in higher education.
- Landgraf, A. J. and Y. Lee (2015). Dimensionality reduction for binary data through the projection of natural parameters. Technical Report 890, Department of Statistics, The Ohio State University.
- Osborne, S. P., Radnor, Z., Kinder, T., & Vidal, I. (2015). The SERVICE Framework: A Public service dominant Approach to Sustainable Public Services. *British Journal of Management*, 26(3), 424-438.
- Osborne, S. P., Radnor, Z., & Strokosch, K. (2016). Co-Production and the Co-Creation of Value in Public Services: A suitable case for treatment?. *Public Management Review*, 18(5), 639-653.

- Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2, 559-572.
- Petruzzellis, L., & Romanazzi, S. (2010). Educational value: how students choose university: Evidence from an Italian university. *International Journal of Educational Management*, 24(2), 139-158.
- Simões, C. & Soares, A. M. (2010) Applying to higher education: information sources and choice factors, *Studies in Higher Education*, 35:4, 371-389.
- Trotter, E., & Roberts, C. A. (2006). Enhancing the early student experience. *Higher Education Research & Development*, 25(4), 371-386.