

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

Giovanni Azzone<sup>a</sup>; Mara Soncin<sup>b\*</sup>

<sup>ab</sup>Department of Management, Economics and Industrial Engineering, Politecnico di Milano, Italy.

\*corresponding author: e. [mara.soncin@polimi.it](mailto:mara.soncin@polimi.it)

This is a post-peer-review, pre-copyedit version of an article published in *Studies in Higher Education*. The final authenticated version is available online at <https://doi.org/10.1080/03075079.2019.1612354>

---

## Abstract

*When investigating students' motivations to enrol in university, a wide range of elements related to the overall student experience should be taken into account. The current study moves from this point to analyse students' choice factors from a survey completed by 27,504 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 proves to be highly heterogeneous across universities, which are selected for different contextual factors even within the same region. From this, a managerial tool is provided to position students population and derive strategic implications. This increases universities' awareness and enables them to better focus on the main served population or to target a different one. Finally, policy considerations are reported.*

**Keywords:** Higher education; university choice; student expectations; principal component analysis; student choice.

---

## 1. Introduction

The level of complexity around and within the higher education (HE) system has increased over time, with major concerns and an open debate about marketisation and growing competition for student recruitment (Jamelske, 2009; Simões & Soares, 2010; Woodall *et al.*, 2014; Nixon *et al.*, 2018). Selecting the university to attend is complex for a population of students that has dramatically widened over time and that more than in the past considers a wide range of factors not only related to teaching and learning quality support (Reay, 1998; Petruzzelli & Romanazzi, 2010; Hemsley-Brown & Oplatka, 2015). Understanding the factors driving students' choice is then relevant to align their expectations to the HE institutional offer. Indeed, a discrepancy between expectations that drive students' choice and their experience of the HE system and institution may bring to a misalignment resulting in higher students' drop-out and, from an organisational point of view, an increasing complexity in managing HE institutions. The current study moves from these considerations in a public policy setting by analysing the dimensions of the university experience that are mostly taken into account by students when choosing the HE institution to attend.

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 choosing a university abroad now consider availability of scholarships as the most important factor, while university rankings have dropped to the last position. Other factors continue maintaining a prominent role in student's choice, such as geographical proximity and university reputation (Briggs and Wilson, 2007; Simões & Soares, 2010). Most of the studies in the field suffer, though, from a major limitation which is related to the method of analysis that do not completely exploit the complexity of students' decision, who may jointly consider multiple factors related to different aspects of their university experience (Reay, 1998). In addition to that, the variability in the distribution of students' choice across institutions has high potential implications that are currently under-investigated, since most of the studies are focused on single institutions. This study empirically analyses 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).

Moving from this, 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 these purposes, a logistic Principal Component Analysis (PCA) has been applied to data on students' choice collected by surveying 27,504 students across 23 public HE Italian

institutions. The originality of the research lies in the creation of a unique dataset that allows to make considerations about students, but also to aggregate students' choice at university level to provide suggestions to managers and policy-makers, by using a methodology that accounts for the multidimensionality of student's decisions. As a remainder of the paper, section 2 reports the related literature, while section 3 explains the conceptual framework; section 4 presents data and methodology; finally sections 5 and 6 respectively present and discuss the results.

## **2. Related literature**

Literature concerning students' choice has been heterogeneous in describing the main factors associated to individuals' decisions. Most of the studies analyse students' choice by means of a number of factors related to the institutions' characteristics, such as the quality of academic services or institutional reputation; some others relate to elements mediating and moderating individuals' choice, such as demographic characteristics or socio-economic factors; finally, elements of interaction between students and institutions, such as geographical proximity or cost sensitivity are considered (Hemsley-Brown & Oplatka, 2015).

Among institutional elements, the university reputation, location and the quality of academic services are the main factors affecting students' choice, together with an element of interaction between student and institution, which is geographical proximity. In their study, Simões and Soares (2010) find that geographical proximity and university reputation are the factors mainly considered by students in a Portuguese institution where 1,641 students have been surveyed. Briggs (2006) analyses a survey to 651 undergraduate students across 6 Scottish universities, finding heterogeneity across institutions and disciplines, but recurrence in the most important factors: academic reputation, proximity and location of the university. Hemsley-Brown (2012) confirms the importance of university reputation in the choice of international students studying in the UK, again followed by the location and by factors related to the quality of teaching. Drewes and Michael (2006) apply a rank ordered logit model to data provided by 27,981 applicants to province universities in Ontario, Canada. They confirm the relevance of the geographical proximity, followed by the amount of resources invested in scholarships and teaching and by the level of non-academic student services. Moving from Briggs (2006), Briggs and Wilson (2007) focus on the effect of costs and information in six Scottish universities by adding a new wave of data (1,400 undergraduates surveyed), finding that course content information plays a prominent role in students' choice. In this strand, other studies analyse students' choice from the point of view of the source of information used. Johnston (2010) finds that the main source of influence is represented by university coaches and staff, while family members and personal sources follow in importance.

A number of studies in the UK context are focused on the effect of an increasing marketisation of the HE system (Nixon *et al.*, 2018), partially realised through an increase in the level of fees required to families from 2012 on. A number of factors are found to be considered by students and families in addition to the contribution level, with particular reference to the course and university reputation (Dunnet *et al.*, 2012), employment possibilities and university location (Burge *et al.*, 2014).

Research in the field somehow explored also the importance of non-academic factors, which relate more to a dimension of student experience. Kallio (1995) stresses the importance of the campus social environment as an influential dimension in the USA, in addition to other academic and work-related variables. Indeed, most of the studies analyse the effect of this dimension *during* the first year of university and not *before* choosing the university itself, studying, for example, the effect of the university social dimension in influencing the probability to continue or drop out at the end of the first year (Wilcox *et al.*, 2005). However, students may also choose a specific university on the basis of personal expectations on the social context and surrounding characteristics, an aspect currently under investigated.

In addition to the characteristics of the institutions, the role of social factors and students' characteristics have been widely investigated with mixed results. In the South-African context, Bonnema and Van der Waldt (2008) analyse answers provided by 716 high school students, finding five clusters of students highly differentiated on the basis of their demographic and socio-economic characteristics, with students coming from less affluent background selecting the university that offers higher opportunities to build a "new life". Wiese *et al.* (2010) study the same national context finding that, in a multiethnic society, the language group is more important than other demographics (like gender) in defining differences across students' choice. In the European context, Mangan *et al.* (2010) apply an ordered logit analysis on data from 1,272 students across the UK to analyse the relationship between social class and university choice; social class does not emerge as a factor directly affecting students' choice, which is indeed mediated by variables related to students' academic career, such as examination grades and the type of high school attended. Sianou-Kyrgiou and Tsiplakides (2011) report opposite findings in the Greek context, where social class strongly affects the way in which the choice of the university is experienced by students.

Apart from social factors, gender is the individual characteristic mostly studied, showing how it may affect students' choice (e.g. Moogan & Baron, 2003). In the Italian context, Cattaneo *et al.* (2017) analyse the effect of the financial crisis on gender differences when choosing university, finding that the crisis may have increased the gender gap through a mechanism that brings men to choose more career-oriented degrees and women to focus more on the current educational experience – a fact that may place them at a higher disadvantage on the labour market than in the past. Also Petruzzelli and Romanazzi (2010) analyse the Italian context collecting 923 questionnaires to study students' perception of value, highlighting the

complexity of students' decisions and the importance of considering the overall student experience.

In their recent review of the literature, Hemsley-Brown and Oplatka (2015) summarise the mixed results emerging from the literature on students' choice, highlighting the heterogeneity of results that makes the higher education student market highly segmented. They also stress the convenience of restricted samples that most of the studies analyse, which are "likely to result in findings which are biased towards the specific strengths of the institution where the study is carried out" (p. 267). Finally, they describe the major attention paid to socio-economic characteristics in the literature to suggest the need for moving towards factors more related to students' lifestyle and behavior.

The current research contributes to the line of literature about students' choice in relation to institutional characteristics and to the factors of interaction between students and institution, by enlarging the set of elements to include the surrounding context at territorial level in terms of costs, quality of life and job opportunities. Moreover, it brings evidence about a national level sample of students, in order to contribute to the multiple institutions' comparison.

### **3. Conceptual framework**

The framework on which this study is grounded relates to the sphere of students' choice. Four main elements interact within the system, as they are summarised in Figure 1. The central actor is the student (1), whose decision upon the HE institution to attend is related to a number of factors that can be traced back to the HE institution's characteristics (2) and to elements of interaction between the student and the HE institution (3). In addition to that, the surrounding context (4) in which the HE institution is located plays an important role.

The choice made by students about the university to attend may be mediated by socio-economic and individual factors, such as socio-economic status, gender or ethnicity (Mangan et al., 2010; Weise et al., 2010; Kyrgiou & Tsiplakides, 2011; Cattaneo et al., 2017). When making a decision about the university, students consider the overall institutional quality, as it may be expressed by the quality of teaching, that of student services or by the national and international reputation of the university (Briggs, 2006; Simões & Soares, 2010; Hemsley-Brown, 2012). In addition to this, elements of interaction between the student and the university are influent along the decisional process. In particular, the geographical proximity highly affects the choice made by students, who are also sensitive to the economic aspects related to enrolment and attendance (i.e. the amount of tuition fees, scholarships or other financial aid) (Briggs, 2006; Drewes & Michael, 2006; Simões & Soares, 2010). A last element to be considered relates to the interaction between the university and the surrounding area. Indeed, universities prepare students for a job market that may be on a large scale as well as local. Hence, the employment rate in the area can influence the choice of students who take a long terms perspective. In addition to this, the overall characteristics of the area

in terms of quality of services and costs of living are relevant components affecting students' decision (Kallio, 1995; Wilcox et al., 2005).

[Figure 1] around here

## 4. Data and methods

### 4.1. 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 yearly 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 detail, questionnaires are filled out by the teaching staff, the administrative body and students. In the wave here considered, 23 HE institutions agreed to join the student survey. The questionnaire is related to the evaluation of administrative services, ranging from the quality of office services to university facilities. In addition to this, the survey for first-year students is enriched by an additional question concerning the factors driving the choice of the university to attend. Details about the formulation are given in Appendix A. Factors 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. Possible answers are those cited in the conceptual framework (Section 3) with one notable exception: to ensure students' anonymity, no personal characteristics are asked to respondents, so that the role of social and individual factors cannot be explored. With reference to all the other factors, options of answer 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 admission 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 is 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. To account for this heterogeneity,

model specifications that consider alternative subsamples are presented in Appendix B. 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] around here

In the survey, the student is not forced to choose only one driver of choice, but can select up to three options, under the assumption that student's choice is multidimensional in nature and, hence, a typology of students' choice can be analysed. Coding results, we record a value of 1 everytime the factor is selected by the student and 0 otherwise, with a maximum of three variables selected per student. Table 2 reports summary statistics about the variables. 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).

[Table 2] around here

#### **4.2. 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 (de Leeuw, 2006; 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) and maximise the variance explained by each component. As in traditional PCA, the selection of the number of principal components is based on different criteria, 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.

The focus of the study, hence, is not only on how frequently a factor was selected, but also on the pattern of selection by each student. The possibility of multiple factors selection is important, as it creates underlying patterns in the data. 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. Moreover, the model used gives the possibility to account for the binary nature of data, but it does not constraint each line of observation to sum up to one. In other terms, it allows for the multiple selection nature of students' answer. These considerations are important to support the methodological choice made throughout the study.

## 5. Results

### 5.1 Baseline results

The number of components from the application of logistic PCA to the 27,504 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 37.2% of the total variance. The second and third components respectively represents the 16.4% and 13.7% of the variance for a total amount of 67.4% of variability explained, which suggest a satisfying fit of the component solution. The loadings for the three principal components are represented in Figure 2. 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. This confirms that students do have a multifaceted approach to university choice. The fact that, among variables, the cost of life in the area reports the highest absolute value highlights the important role played by the surrounding conditions when university is selected. The second component is instead represented by the contraposition of two variables, which have the highest loading values: 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 get 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. As examples of possible maps to be created at institutional level, Figure 3 reports students' distribution along components that have a remarkable contrast between variables: the second (x axis) and the third components (y axis). As a first insight, it is possible to investigate students' distribution within institutions. Indeed, each of the maps shows the density of students' concentration along dimensions. This way, polarisation of students' groups emerge, by allowing to build clusters of students moved by similar motivations. For instance, the map for University M highlights two main groups of students: those selecting the HE institution for the easy access and those who valued the proximity to the home town (red areas, with a higher density of individuals). On top of this, the graphical representation allows to visualise additional groups smaller in size, but relevant to cluster "marginal" groups of students whose drivers of choice differ from the majority. In the former example, students selecting University M for its reputation (yellow spot on the bottom) are a striking case, as well as students more prone to consider job market opportunities (yellow spots on the left). A similar pattern is visible across HE institutions, showing that even if some elements approximate the majority of students' choice more than others, different groups of students are present within universities.

In addition to this, the richness of the dataset enables to compare results across institutions. The comparison provides the possibility to highlight how the heterogeneity emerged within institutions is even more detectable in the cross-university comparison. Examples for three HE 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 territorial context, factors of choice may profoundly differ. For instance, a high concentration of students selects University M because of its proximity to home and ease of access (red areas in the map), as previously mentioned. University N is

instead mainly targeted for its reputation and for the job opportunities in the region (red areas in the bottom left). This dimension of choice concerns indeed a broad driver of selection for a HE institution, which includes the university itself and the territorial context in which the institution is located. In this sense, it relates to that array of characteristics that are part of the *overall experience* (in this case, related to the employability dimension) that students do consider in their choice. What is worth to be noticed is that a number of students select University N because of the territorial job opportunities, while this driver is not relevant for University M and it is much less evident for University S, despite the fact that institutions are located in the same regional context and, for University N and S, in the same city. This fact stresses the high heterogeneity emerging from the cross-institutional comparison, highlighting the fact that students choose university for a wide array of factors partially related to contextual conditions that may be in common across institutions but differentiated in students' perspectives.

[Figure 2 and 3] around here

## 5.2 Robustness checks

To check for the robustness of findings, two alternative approaches have been applied. The first consists of using the same methodology (a logistic PCA) with an additional number of components, in order to investigate the differences in components' composition. The second consists of applying a PCA to the data "weighted" for the number of factors selected by each student. As mentioned in the previous section, students can select up to three drivers of university choice; hence, each observation has been equally weighted by the number of other factors selected per student (hence, for a student selecting two motivations, a value of 0.5 is attributed to each of the selected variables).

Analysing the four components logistic PCA, the first dimension still emerges as the average of the different factors, confirming the high variability of data. The four variables opposing one another in the three-component model (i.e. proximity, job opportunities in the area, reputation and ease of access) are still present in the four-components specification. As reported in Columns 2-5 of Table 3, proximity is mainly opposed to quality of student services in the second component, while job opportunities in the area is opposed to economic factors in the third component. Reputation and ease of access are confirmed as contrasting one another in the fourth component, as it is in the baseline model.

When analysing the "weighted" model, the three-component specification is able to explain the 68.8% of the variability, similarly to the baseline model. Moreover, as shown in Columns 6-8 of Table 3, results confirm the relevance of variables originally specified, but differently mixed. Indeed, the first component now shows the contrast between reputation and

proximity; the second component reports reputation and job opportunities in the area; the third component mainly reports job opportunities in the area and ease of access.

Overall considered, the variables emerged in the baseline model are indeed confirmed in alternative specifications.

[Table 3] around here

## 6. 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 the university to attend, and that a clear understanding of these dimensions align students and HE institutions' perspective, increasing the overall social utility. Applying a logistic PCA to a dataset of 27,504 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. The four dimensions emerged as relevant partially confirms findings in the literature for the factors related to the institutions, like university reputation or proximity to the home town (Briggs and Wilson, 2007; Simões & Soares, 2010; Hemsley-Brown & Oplatka, 2015). The relevance of the variable related to the ease of access provides an interesting insight about the different facets of students' choice. In addition, the need of keeping a broader view when considering factors that may affect students' choice is highlighted by the importance of the job opportunities in the region. This variable cannot be directly related to the specific institution, but indeed concerns the territorial context and, somehow, relates to a broad definition of student experience where the student interact with the institution and with the territorial context as well.

Looking at the graphical maps (Figure 3) a great variation in students' distribution across and within HE institutions emerges. This kind of evidence may be used by institutions at managerial level 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. 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 to a higher "internal" personalisation of services offered to students with different attitudes and motivations. Indeed, the ability to identify from the

very beginning the drivers that characterise students' choice enables the provision of services that are more targeted and aligned to the expectations and needs of the students served. In this perspective, the HE institution characterised, for instance, by a certain number of students enrolling because of the job opportunities in the region may increase the personalisation of job placement services even from the earlier stages of the university path. At policy level, information provided by the analysis of students' choice 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 centre of value creation. In this sense, the possibility to compare different institutions across the country allows for the identification of similarities and contrasts across universities.

In this study, we argue that the alignment of perspectives between students and HE institutions is of central importance in the generation of public value. This is in line with the approach proposed by Osborne *et al.* (2015), according to whom it is necessary to rethink public services placing the user at the centre of the *service experience*. 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). First year students represent a case in point in this respect, given the tense expectations they have when they enter tertiary education with respect to their future university experience (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, defined as 'a coherent whole that requires a facilitative campus environment and student effort' (Petruzelli & Romanazzi, 2010; p. 143).

As a future development of the study, following students over time and analysing the relationship between university choice and subsequent academic path or drop out probability would add a further layer of analysis to the current investigation. Moreover, students' personal information (such as previous career or socio-economic status) is not collected at current stage 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. Yet, the current analysis provide highly usable insights on the factors driving students choice from a cross-institutional and multi-dimensional point of view, showing the heterogeneity and communalities behind students' decision of the institution to attend. Finally, we acknowledge that the selection of a specific methodology implies benefits and limitations. Alternative models like factor analysis or structural equation modelling could have shed a different light on data analysis. Although, we firmly believe that the features of this model, which specifically accounts for the binary nature of data and supports an explorative approach to dimensionality reduction, make the methodology employed the most

appropriate for our purposes. Despite this, we will consider the possibility to explore other methods in similar applications in the future.

## References

- ANVUR. (2016). Rapporto sullo stato del sistema universitario e della ricerca 2016.
- Bonnema, J., & Van der Wal, D. L. R. (2008). Information and source preferences of a student market in higher education. *International journal of educational management*, 22(4), 314-327.
- Briggs, S. (2006). An exploratory study of the factors influencing undergraduate student choice: the case of higher education in Scotland. *Studies in Higher Education*, 31(6), 705-722.
- 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.
- Burge, P., Kim, C. W., Rohr, C., Frearson, M., & Guerin, B. (2014). Understanding the Impact of Differential University Fees in England. Rand Europe.
- Cattaneo, M., Horta, H., Malighetti, P., Meoli, M., & Paleari, S. (2017). Effects of the financial crisis on university choice by gender. *Higher Education*, 74(5), 775-798.
- De Leeuw, J. (2006). Principal component analysis of binary data by iterated singular value decomposition. *Computational Statistics & Data Analysis*, 50(1), 21-39.
- Drewes, T., & Michael, C. (2006). How do students choose a university?: an analysis of applications to universities in Ontario, Canada. *Research in Higher Education*, 47(7), 781-800.
- Dunnett, A., Moorhouse, J., Walsh, C., & Barry, C. (2012). Choosing a university: A conjoint analysis of the impact of higher fees on students applying for university in 2012. *Tertiary Education and Management*, 18(3), 199-220.
- Hemsley-Brown, J. (2012). 'The best education in the world': reality, repetition or cliché? International students' reasons for choosing an English university. *Studies in Higher Education*, 37(8), 1005-1022.

- Hemsley-Brown, J., & Oplatka, I. (2015). University choice: what do we know, what don't we know and what do we still need to find out?. *International Journal of Educational Management*, 29(3), 254-274.
- Johnston, T. C. (2010). Who and what influences choice of university? Student and university perceptions. *American Journal of Business Education*, 3(10), 15-24.
- Kallio, R. E. (1995). Factors influencing the college choice decisions of graduate students. *Research in Higher Education*, 36(1), 109-124.
- 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.
- Mangan, J., Hughes, A., Davies, P., & Slack, K. (2010). Fair access, achievement and geography: explaining the association between social class and students' choice of university. *Studies in Higher Education*, 35(3), 335-350.
- Moogan, Y. J., & Baron, S. (2003). An analysis of student characteristics within the student decision making process. *Journal of further and Higher Education*, 27(3), 271-287.
- Nixon, E., Scullion, R., & Hearn, R. (2018). Her majesty the student: marketised higher education and the narcissistic (dis)satisfactions of the student-consumer. *Studies in Higher Education*, 43(6), 927-943.
- 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.
- Petrizzellis, L., & Romanazzi, S. (2010). Educational value: how students choose university: Evidence from an Italian university. *International Journal of Educational Management*, 24(2), 139-158.
- Reay, D. (1998). 'Always knowing' and 'never being sure': familial and institutional habituses and higher education choice. *Journal of Education Policy*, 13(4), 519-529.
- Sianou-Kyrgiou, E., & Tsiplakides, I. (2011). Similar performance, but different choices: social class and higher education choice in Greece. *Studies in Higher Education*, 36(1), 89-102.

- 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.
- Wiese, M., Jordaan, Y., & Van Heerden, C. H. (2010). The role of demographics in students' selection of higher education institutions. *Professional Accountant*, 10(1), 150-163.
- Wilcox, P., Winn, S., & Fyvie-Gauld, M. (2005). 'It was nothing to do with the university, it was just the people': the role of social support in the first-year experience of higher education. *Studies in higher education*, 30(6), 707-722.
- Woodall, T., Hiller, A., & Resnick, S. (2014). Making sense of higher education: Students as consumers and the value of the university experience. *Studies in Higher Education*, 39(1), 48-67.

**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          | 1,494 | 52%           | N          | 1,028  | 21%           |
| C          | 2,355 | 53%           | O          | 1,395  | 8%            |
| D          | 336   | 4%            | P          | 191    | 5%            |
| E          | 1,774 | 63%           | Q          | 157    | 5%            |
| F          | 245   | 33%           | R          | 417    | 15%           |
| G          | 1,427 | 19%           | S          | 1,155  | 9%            |
| H          | 1,191 | 11%           | T          | 2,920  | 76%           |
| I          | 980   | 16%           | U          | 407    | 11%           |
| J          | 340   | 7%            | W          | 694    | 16%           |
| K          | 655   | 19%           | X          | 5,492  | 70%           |
| L          | 20    | 14%           | Total      | 27,504 |               |

Note: The response rate is calculated as the number of respondents over the total number of first-year students.



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

| Variable                      | N      | Percentage of students selecting the variable |
|-------------------------------|--------|---|
| Economic factors              | 4,195  | 15%   |
| Reputation                    | 9,445  | 34%   |
| Ease of access                | 5,269  | 19%   |
| Quality of student services   | 3,160  | 11%   |
| Proximity to home             | 12,113 | 44%   |
| Quality of life               | 2,763  | 10%   |
| Job opportunities in the area | 6,169  | 22%   |
| Cost of life                  | 1,211  | 4%  |

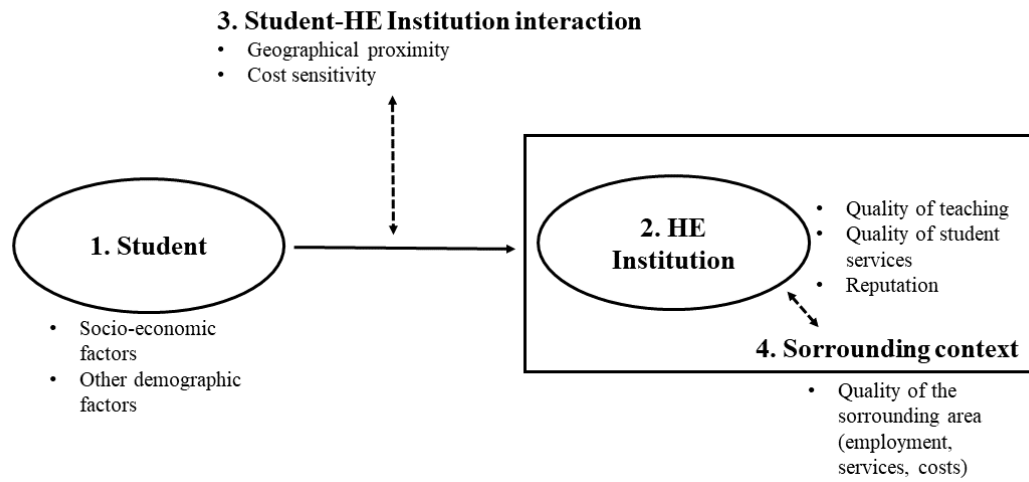
Note: Sum of percentages is greater than 100% given the possibility to select more than one variables per respondent.

**Table 3. Factor loadings across alternative models.**

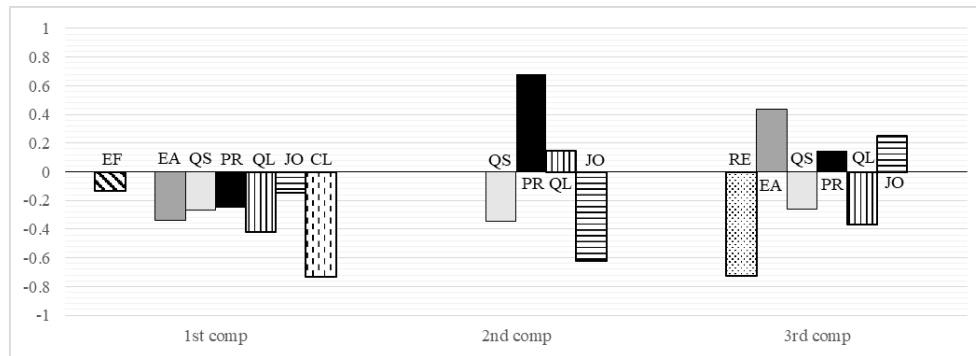
|                               | 4 components Logistic PCA |          |          |          | 3 components "weighted" PCA |          |          |
|-------------------------------|---------------------------|----------|----------|----------|-----------------------------|----------|----------|
|                               | 1st comp                  | 2nd comp | 3rd comp | 4th comp | 1st comp                    | 2nd comp | 3rd comp |
| Economic factors              | -0.23                     |          | -0.677   |          |                             |          | 0.181    |
| Reputation                    |                           | 0.257    | -0.269   | 0.633    | -0.434                      | 0.744    | -0.253   |
| Ease of access                | -0.165                    | 0.252    | -0.191   | -0.614   |                             | -0.184   | 0.673    |
| Quality of student services   | -0.177                    | 0.658    |          |          |                             |          |          |
| Proximity to home             | -0.344                    | -0.634   |          |          | 0.873                       | 0.220    | -0.238   |
| Quality of life               | -0.504                    |          | 0.102    | 0.444    |                             |          |          |
| Job opportunities in the area | -0.305                    | 0.174    | 0.642    |          | -0.205                      | -0.601   | -0.623   |
| Cost of life                  | -0.650                    |          |          | -0.152   |                             |          |          |

Note: Factors loadings reported. Varimax rotation applied.

Figure 1. Conceptual framework.

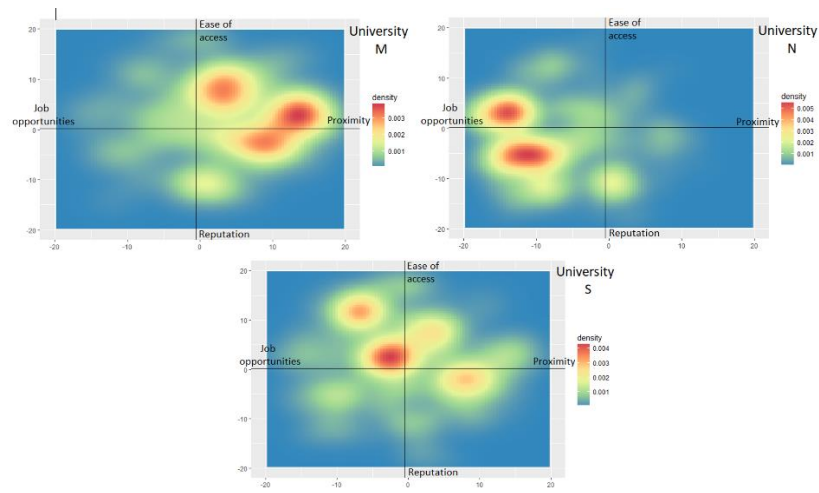


**Figure 2. Factor loadings along the three principal components.**



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 3. 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.*

## Appendix A

Questionnaire.

Here is presented the formulation of the question related to the factors of choice as submitted to the students. Each respondent could select from a minimum of one to a maximum of three options among those proposed in the survey.

**Question:** *Choose, among the following, the aspects that most influenced your University choice (max 3 options)*

**Options:**

- *Economics aspects (University fees, scholarships)*
- *University reputation*
- *Admission requirements (lack of the entrance exam, prerequisites)*
- *Students' services (information availability, facilities, help and guidance)*
- *Proximity to the hometown*
- *Quality of life (public services, amenities, local area development)*
- *Job opportunities in the region (average salary, employment rate)*
- *Cost of living*

## Appendix B

### Alternative sample specification.

Given the heterogeneity in the response rate across HE institutions, two alternative sample specifications have been verified considering different subgroups of observation in order to test for the robustness of findings.

In the first specification, only HE institutions where at least 10% of students answered have been considered, for a total amount of 23,581 students across 16 institutions. Results are reported in columns 2-4 and show how the first component is still a combination of the different factors, with the cost of life assuming an even more prominent role with respect to the baseline model. The second component confirms the importance of the proximity to home as opposed to the quality of student services. Finally, the third component also confirms the relevance of the reputation as opposed to the job opportunities in the region. In this specification, students who choose university because of its reputation also consider the quality of life in the area where the university is located. Hence, the main drivers of choice (three out of four main variables) are confirmed in the specification that consider this subsample of institutions.

In the second specification, only HE institutions where at least 15% of students answered have been selected, for a total amount of 19,481 students across 12 institutions. Results are reported in columns 5-7 and strongly confirm the baseline model. Indeed, the first component is a heterogeneous combination of the different possible factors. The second component is instead given by the contraposition of proximity to home and job opportunities in region. Finally, the third component is given by the reputational dimension as opposed to the ease of access.

**Table B1. Factor loadings across models using alternative subsamples.**

| Variable                      | 3 components Logistic PCA –<br>Respondents > 10% |          |          | 3 components Logistic PCA –<br>Respondents > 15% |          |          |
|-------------------------------|--|----------|----------|--|----------|----------|
|                               | 1st comp   | 2nd comp | 3rd comp | 1st comp   | 2nd comp | 3rd comp |
| Economic factors              | -0.135   |          |          | -0.13  |          |          |
| Reputation                    | 0.152  | -0.265   | -0.634   | -0.26  |          | -0.714   |
| Ease of access                | -0.143   |          |          | -0.248   |          | 0.558    |
| Quality of student services   | -0.243   | -0.624   | -0.165   | -0.158   |          |          |
| Proximity to home             | -0.325   | 0.625    |          | -0.128   | 0.691    | 0.229    |
| Quality of life               | -0.404   |          | -0.55    | -0.468   | 0.161    | -0.249   |
| Job opportunities in the area | -0.256   | -0.382   | 0.504    | -0.15  | -0.697   | 0.21     |
| Cost of life                  | -0.738   |          |          | -0.756   |          | 0.137    |
| # of observations             | N=23,581   |          |          | N=19,481   |          |          |

Note: Factors loadings reported. Varimax rotation applied.