International Journal of Science Academic Research
Vol. 05, Issue 03, pp.7220-7229, March, 2024
Available online at http://www.scienceijsar.com

ISSN: 2582-6425

Research Article

EFFECTS OF GAP-YEAR ON STUDENTS' TIME TO DEGREE

***Mattia Fasano**

University of Calabria, Italy

Received 17th January 2024; **Accepted** 25th February 2024; **Published online** 29th March 2024

Abstract

Previous research work analyses the gap-year and its effect on students. In this study the main idea is to link the gap-year with the students' performance. As a measurement of this performance, the time spent from the students to reach their degree is used. This study uses an Italian dataset of students who took their degree from 2010 to 2015, in five-year faculties. The study is split in two parts: in the first one we analyse the duration of the studies, from the enrolment to the degree; in the second one we use a dummy variable to differentiate students who reach their degree in five years or less and students who had an extensions of their studies. In the end, our study finds that gap-year students spent less time to reach their degree.

Keywords: Time-to-degree, Italian students, Gap year, Students' performance.

INTRODUCTION

The goal of this study is to show whether having one or more gap years can affects students' performance. During this work, performance is measured through the "time to degree", that can be seen as the time spent by a student from the enrolment in the university until the graduation. More specifically, the study is focused on the possibility that students will take less time to complete their studies after having the gap year. In fact, the previous research on the gap year were focused on the causes that bring the student to decide to have or not this gap year and for this reason the decision of analyse the consequences, instead of the causes, seems to be an interesting focal point. The ultimate goal of this study is to demonstrate that it is possible that a student who had a gap year is motivated to reach the degree faster, and so, to spent less time in the university due by the fact that they are "in a hurry" to make up the time that was spent in a different way. Nowadays this topic is becoming more important because the number of students who had a gap year is rising every academic year, and if some years ago this number was very low, now it starts to become more significant. The main idea is to take in account Italian students from a specific dataset: the Almalaurea data[1]. Almalaurea is an Italian firm that collects data from all the Italian university to analyse these data and create some statistics that once per year are showed during the annual conference. Almalaurea gives the possibility to who present a request to work with his own dataset as well. For what concern the data analysed in this study, it contains information on the five-year degree students from all Italy. This means that there is not information on the bachelor's students (three-years degree) as well as postgraduate's one (two-years degree), but just about students that choose some specific faculties, like law, veterinary, engineer and some others. The information on these students are about their socio-economic background, on their families, where they born, live and study and so on.

Corresponding Author:** ***Mattia Fasano
University of Calabria, Italy

[1]https://www.almalaurea.it/en

The main information that are taken from this dataset is the time spent to reach the degree. With this information is measured the performance of the students. It is important to analyse their performance because after the university these students will join the job market and, in that moment, they will represent the human capital. Since the human capital is a very important index for the growth of a country, is important to analyse the education path of the students, to understand what level of human capital will be available in the job market, and as a consequence, how much a country can grows up. Since to make a correct analysis it is important that the students reached their degree, in this study are take in account just students who completed their university's studies. For this reason, the analysis takes in account only students who took their degree from 2010 to 2015. In the second chapter, after this introduction, we provide a literature review; this chapter is split in two parts: the first one is focused on the "time to degree", the second one highlights the "gap year". Below, in the third chapter there is the main body of this study; this part is divided in three sub-chapter: the presentation of the methodology, the descriptive statistics, and the results obtained using the STATA software and highlighted using graphs. Finally, we conclude.

LITERATURE REVIEW

Time to degree

Extending the undergraduate studies beyond the standard degree period is not so uncommon. In fact, since in the Italian university the students can choose when they want to do an exam, there is the possibility that some students complete their degree beyond the standard degree time; in this way these students have an extension of studies. As is showed later, different studies tried to analyze it (see *inter alia* Aina and Pastore 2012, Kurzweil *et al.* 2015) and a decrease of this phenomenon was observed in 2001 when in Italy the higher education change into the new system "three plus two", that consists of a bachelor degree of three years, plus a master degree of two years.

First, we focus on the consequences of extending undergraduate studies beyond the standard degree period and how this can affect the job satisfaction. Brodaty *et al.* (2008) points out that students who extended their studies had negative effects on their job. More specific, for an additional academic year to reach the degree there is a 9% income reduction during the first years of the working life. On the other hand, Aina and Casalone (2011) find that only with a large delay of at least two years there is a reduction on the wage.

Moreover, Aina and Pastore (2012), using the same Almalaurea dataset show that have an extension of studies has another consequence: for each year of delay there is an increase of 2% of the probability of doing a kind of job that is not linked with the degree. To reach this result Aina and Pastore (2012) estimate a probit model using as dependent variable a binary variable with value 1 for overeducated students and 0 otherwise; among the independent variable there is one variable that highlight the year of delay in reaching the degree. Just later, other researchers, as Pastore (2009), started to work on the causes that bring the students to an extension of studies. This happen because initially the idea was that take more years was due to student's behaviour or choices. Recent studies, that are pointed out later on, instead, conclude that this phenomenon is not caused just because of the students but there are other reasons that are highlighted after. Of course, how much effort the students put during their studies, that can be measured through the attendance or the usage of library or in others way, is very important in this topic, in fact, according to Alma Laurea, a firm that collects student's data from every public university in Italy, students who attend at least the 75% of their courses have a higher probability to spend less time in order to reach their degree than other students. Moreover, also the socioeconomic background of the students affects what they do: Bound *et al.* (2012) pointed out a difference in term of time to degree between students from above and below the median of income distribution. More precisely, students with the lower income had the bigger increases in time to degree. To reach this conclusion Bound *et al.* (2012) use the data from the National Longitudinal Study of 1972 (NLS72) and the National Educational Longitudinal Study of 1988 (NELS:88), to analyse the real income level of the family, divided in six blocks, to measure the ability to pay the tuition fees.

At the same time, it is important to take in account the other factors that can affect the time spent to reach the degree. For example, a well-organized university with efficient services can help the students to be faster, as Kurzweil *et al.* (2015) shows. This study point out that an increase of the number of enrollment, that was not followed by an increase of resources available for students, brings a decrease of students performance and consequentially an increase of the time spent to reach the degree.

Considering that the aim of acquiring a degree is to enter the job market and pursue a career, could be of particular importance to study the effect of the job market to students' performance and estimate the opportunity cost of studying. About this topic, Becker (2001) provides an empirical evidence on the "parking lot hypothesis". Becker (2001) shows that a greater unemployment rate decreases the opportunity cost of studies and so it brings an increase of enrollment rate in university rather than a direct enter in the job market. In fact, in Italy in the zones where the unemployment rate is higher there is a greater number of students who took more years to reach the degree. Using the previous research on this topic some researchers, like Pastore (2009) and Becker (2006), developed different economics models. One of these models is the "human capital model" presented by Pastore (2009); this model follows from the human capital theory of Becker (1962). The researcher tried to find the causes of the extension of studies, analyzing all the factors that can either reduce the student's performance, in terms of grades and time, or increase the direct and indirect cost, that are represented by the tuition fees and the other cost linked with the student's life. Among these factors, there are: very low skilled worker demand and university inefficiency. Moreover, this model can be used to rationalize the effect of being late in taking the degree. In fact, Pastore (2009) highlight that, if take more years can slow down the enter in the job market, and so the work experience, on the other hand this delay can be due to a higher level of training and formation that can increase the value of the human capital. For this reason, is very important to understand "how" this time has been spent.

Last but not least, another model is the "Job Search model" (Becker, 2006). This model is focused on the fact that the investment to pay the tuition fees and enroll in the university is not just a monetary investment, but a time investment as well. This project takes in account two types of people: those with high school degree and those with university degree. The formers have a decision that must be taken: enroll in the university, or directly enter the market job. This choice, according to Becker (2006), is influenced by different factors. Among these factors there are: job opportunity and university path's difficulty. If the choice is to search for a job this means an increase of unskilled worker and, probably, a decrease of the average "time to degree".

After the analysis of this topic, it is also important to search eventually solutions to the students' problem, that in this case is the fact that the students take their degree after an extension of the required time and this means start to search for a job at older age; which, in turn, means have less chances to get hired (Lahey, 2008). Some solutions to solve this extension of studies' problem were already pointed out but are not easy to be implemented because they need to act at the same time on more aspects. Some of these aspects are highlighted in some studies, like Kurzweil *et al.* (2015), that is pointed out later. These aspects are: the enroll mechanism in the university, didactical organization and a network between university and job market. It is rather important that during the secondary education proper information about higher education is distributed to students to ensure that the students who wish to pursue a degree in higher education have all the necessary information to make the right decision. We might argue that it is rather important to minimize the time between receiving a degree and getting a job. Until now, only private universities in Italy offer internship opportunities to their students and this is something that could potentially benefit students who study at public universities as well.

On the practice point of view, some universities started some experiments, trying to reduce the time to degree and the abandon rate. One of the experiments proposed is the one of Kurzweil *et al.* (2015) about the Georgia State University. The direction board of the university hired 42 news students' advisors, creating a new web site to help them. Through this website, the university created a database of factors that can

affect the probability of the students to extend their studies. There are more than 700 factors in this database, and they indicate if the students did something that brings them to join the group of students who are at risk of take more years to reach the degree. Since this website, named "GPS Advising" (Graduation Progress Success Advising), was in function there were some positive results. According to GPS Advising at Georgia State University[2] report of 2013 the number of degrees was the higher of ever in the Georgia State University history and, in the same period, the 64% of the students of the second year were in the not-a-risk group, according to the database system. This experiment shows that an increase of services and organization inside the university can effectively help the students' performance (GPS Advising at Georgia State University report, 2013).

Gap year

Before to highlight the main part of this topic, it is important to analyse what it means for a student the transition from high school to university. According to Savickas (2005) this transition occurs just in a period in which young people start to join the adulthood. So, it is a very important period of the life, also because the younger start to gain more confidence. In fact, Savickas (2005) using the "Career Style Interview" (Savickas, 1989) that is an interview designed to understand the personality type and the career adaptability, shows that the moment in which students join adulthood is an opportunity for them to develop their own identities and self-beliefs. This personal development is more important in this period because there can be some uncertainty around what the students have to do after the high school and, as is pointed out after, the gap year can be helpful also in this topic (Savickas, 2005).

After this introduction of what represent the transition from high school to university, it is possible to start to point out deeply what is a "gap year". First of all, according to Martin (2010), the decision to take this period, can be due by different reasons: having some rest, starting a part-time job, doing some volunteering. Nowadays, the number of students taking a gap year is increasing. One example of this is the studies of Krause *et al.* (2005) that shows how in Australia this number has grown from 4% to 11% in 30 years. To do that the researchers analyse a survey for the first-year students in Australian universities. This survey is named "First Year Experience Questionnaire" and is conducted at five-year interval. In the survey of 2004, that is the one used in this study, the researchers were able to reach the 28% of response rate, finding information about students' expectation, occupational purpose, enrolling reasons, and others information.

The term gap year is commonly used to describe the discontinuity between secondary and higher education. A gap year can happen due to various reasons, such as the willingness to get a job after pursuing secondary education or the need of time to make a decision on the field into which to pursue further studies (Martin, 2010). In Italy, this phenomenon is not so common; in the AlmaLaurea[3] dataset used in this project, just the 6.25% of the students had one or more gap years. The literature has been rather limited in this area. For example, Coetzee et Bester (2009) show that students who had that "gap year" felt an increase of their own confidence for their future. In the meanwhile, according to the same researchers, this period can create a sense of safety that can help young people in their growth. Moreover, Martin (2010) analyses the effect of the gap year on students' performance while Parker *et al.* (2015) analyses the same effect but on a psychological point of view. According to Martin (2010), it is possible to point out some benefits of the gap-year, as personal and cultural development. In these projects, Martin (2010) uses two studies and Parker *et al.* (2015) use two studies as well. In the latter, the two analysis are different because of the place: one analyses Finland students and the other one Australian students; the goal of this study is to understand whether students who had the "gap year" would have had different outcomes if they had enrolled university directly. Instead, Martin (2010) points out two analysis in which the first one analyses how a lower motivation and uncertainty can bring a student to choose the gap year solution; in the second one, it is showed how the participation to the gap year can predicts a bigger academic motivation. Both these effects are statistically significant. To reach these results, Martin (2010) uses an 11 factors index, called MES-HS. This index assesses motivation through cognitive and behavioural dimensions, and it takes in account the students' development and growth. For each factor there was seven levels reply, from "strongly disagree" to "strongly agree". The statistical analysis is conducted using two models: Confirmatory factor analysis (CFA) and Structural equation modelling (SEM). Both of these models are estimated by the Maximum Likelihood. On the other hand, Parker *et al.* (2015) use a different strategy: to analyse the developmental trajectories of cognitions related to the career goals of the students in the dataset, Parker *et al.* (2015) use a series of growth curve models for all the variables of interest. Both these studies, although the different methodology, bring to the same conclusion: there is a little difference in life and career satisfaction between students who had a "gap year" and students who did not. After the discussion above, it is evident that research, so far, has pointed out that there is a correlation between the gap year and student performance. For this reason, can be important to investigate in a more specific way how this choice can shape the students' pathway.

METHODOLOGY

Models

This study examines the relationship between the gap year and the time to degree of Italian students in university and whether the number of gap years impacts the time to degree as well. This study uses four different methods: linear regression model, Kruskal-Wallis model, Chi-squared test and logistic regression. The idea to use these methods come from other studies that used the same methodology to explore the same field, as Mortada *et al.* (2018) and Hamir (2011). The former study tries to understand the factor that affect the students' performance, meanwhile the latter is more focused on the time to degree. Moreover, using four methods helps to have at least two methods per each dependent variable, since in this project is used first a continuous variable and then a dummy variable. In this way, if the second test shows a similar result to the first one, this can be seen as a confirmation. The methods are split in two different sections, according to the dependent variable that is used. In the first section the dependent variable shows the time spent by the student to take the degree ("duration"). The two models of this section are the Linear regression model and Kruskal-Wallis model. The choice to use these two

[2]https://oie.gsu.edu/files/2014/04/Advisement-GPS.pdf
[3]https://www.almalaurea.it/en

methods is due to the fact that the first one is used by Mortada *et al.* (2018) to analyse the students' performance, that is something similar to this study; meanwhile, the second one is most used by Hamir (2011) to proof a relationship with the time to degree, and this project wants to analyse it as well. In the second section the dependent variable is different; instead of a continuous variable there is a dummy variable. This variable could have two values: 1 if the student spent more than five years to complete their degree and 0 otherwise ("outtime"). The choice of five years is due to the fact that in Italy the legal time to accomplish the degree is five years, this means that a student has to spend at least this time. But, since the students can choose when do the exams, sometimes can happen that they spend more time. From now on, in this project, the sentence "out of time" is used to indicate students who spent more than five years.

The two tests used in this second part are the Chi-squared test and the Logistic regression test. The Chi-squared test was chosen because is very easy to set up through STATA and the logic behind the process can be helpful to highlight this topic as well. For what concern the logistic regression, it is the second method used by Hamir (2011), and since that paper is closer to this project, the idea to use the same methodology seems like a good strategy. To sum up, these four methods are chosen because of two main reasons: the good fitting to the dependent variables that are used and the fact that were already used for similar studies.

Linear Regression: The first one is a simply linear regression, using the time spent by students to reach their degree ("duration") as dependent variable and some of the variables, that are showed in the next paragraph, as independent variable.

$$\begin{aligned} duration = {} & \beta_1 HSMark + \beta_2 y\_gap + \beta_3 age + \gamma_1 male \\ & + \beta_4 exam\_mark + \beta_5 MotherEduc \\ & + \beta_6 FatherEduc + \gamma_2 d\_geoplace1 \\ & + \gamma_3 d\_geoplace2 + \gamma_4 d\_geoplace3 \\ & + \gamma_5 d\_geoplace4 + \gamma_6 d\_geoplace5 \\ & + \beta_7 UniFeedback + \beta_8 Old\_Uni \\ & + \beta_9 DegreePoint + \beta_{10} Uni\_Work \end{aligned}$$

In this regression γ is used to identify the coefficients for the dummy variables.

Kruskal-Wallis test: In some similar works, the ANOVA model was used. This model is a one-way analysis of variance and it is used to determine if the mean of the dependent variable is the same in some unrelated, independent groups. The model requires that the groups used as independent variables are normally distributed. Unfortunately, carrying on the Shapiro-Wilk test, the independent variable ("y_gap") is not normally distributed, so it is not possible to use the ANOVA model in this data. For this reason, the best choice was to use the non-parametric version of that model: the Kruskal-Wallis test. This test is rank-based and can be used to determine if two or more groups, of an independent variable, are statistically significant different on a continuous dependent variable. This test statistic, H, is defined as:

$$H = \frac{1}{S^2}\left\{\sum_{j=1}^{m}\frac{R_j^2}{n_j} - \frac{n(n+1)^2}{4}\right\}$$

Where, n is the overall sample size, R_j denote the sum of the ranks[4] for the jth sample and S^2 is equal to:

$$S^2 = \frac{1}{n-1}\left\{\sum_{all\ ranks} R(X_{ji})^2 - \frac{n(n+1)^2}{4}\right\}$$

The test follows a χ^2 distribution with $m-1$ degrees of freedom.

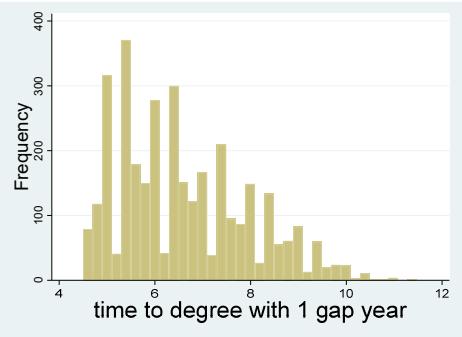
The independent variable used in this model represents the number of years in which the student did not enrol in the university ("y_gap"). So, every year represent one group. Meanwhile, the dependent variable points out the students' time to degree ("duration").

Before to carry on this test is necessary to satisfy some assumptions that are highlighted in the studies of Chan and Walmsley (1997):

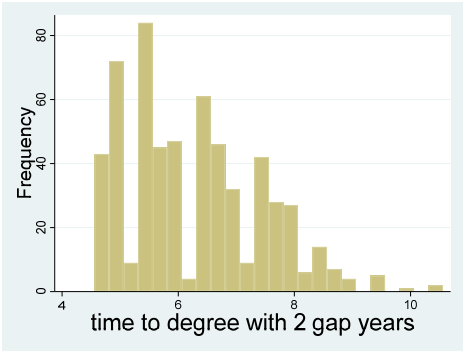
1. The two variables should be measured on an ordinal scale or a continuous scale.
2. The independent variable should consist of two or more categorical, independent groups. Examples of categorical variables include gender (e.g. two groups: male and female).
3. There should be independence of observations, which means that there is no relationship between the observations in each group or between the groups themselves.
4. In order to know how to interpret the results from a Kruskal-Wallis H test, we need to determine whether the distributions in each group have the same shape (which also means the same variability).

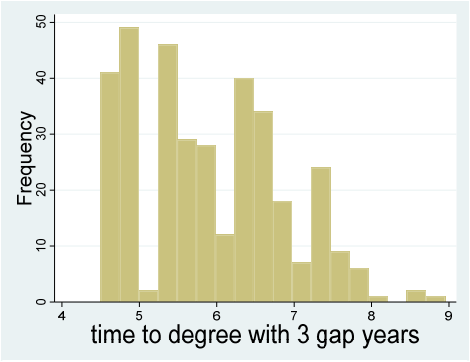
As was pointed out before, the dependent variable is continuous and the independent one is an ordinal scale, so the first assumption is confirmed. Moreover, the independent variable consists in seven categorical and independent groups and, since each student belongs to only one group, this means that the same student cannot be in more groups and so the observations are independent because there is no relationship among different students. With these two sentences, assumptions 2 and 3 are satisfied. To satisfy the last assumption, it is possible to use the histogram of frequencies to show the shape. The Figure 1 below point out that the shapes of the groups are quite the same.

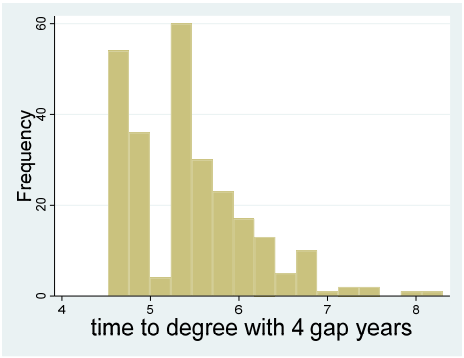
Figure 1. Histogram of frequencies per gap years



[4]Sorting the data for all groups into ascending order and assign ranks to the sorter data. Then add up the different ranks for each group.







These graphs point out how long it takes for the students to reach their degree; in each graph there are students with different number of gap year. On the Y axis there is the number of students (frequency), meanwhile on the X axis there is the variable that shows the two-digit rounded time to degree ("dur"). So, the graphs show how many students (Y-axis) take their degree in that period of time (X-axis).

As pointed out before all the assumptions are satisfied, so it is possible to use the Kruskal-Wallis test.

Chi-square test: The chi-square test for independence, also called Pearson's chi-square, is used to test for a relationship between two categorical variables. The null hypothesis of the test is that the variables are independent. The test will compute the expected theoretical frequency of the two variables, supposing that are independent and then it will compare it with the frequency observed in the data. If the latter fit the model, this will prove that the null hypothesis is correct.

The test statistic is defined as:

$$\chi^2 = \sum \frac{(observed - expected)^2}{expected}$$

In this project, the test is used to check for the correlation between two categorical variables. One shows if the student spent more than five years to reach their degree ("outtime"), the second points out if the student had a gap year or not ("gapy").
It is important to check some assumptions, as pointed out from Wright (2010):

1. The two variables should be measured at an ordinal or nominal level.
2. The two variables should consist of two or more categorical, independent groups.

In this case, the two variables used are measured at a nominal level and since the value they take is either 0 or 1, these variables are split in two categorical groups that are independent between them.

Logistic Regression with prediction: Since in this second part of the analysis the dependent variable is a dummy, it is possible to use the logistic regression model. Moreover, since in other researches, as Hamir (2011), this kind of regression is used to analyse the time to degree, can be helpful use the same methodology to explore the same concept. Using the variable that shows if students spent more than five years ("outtime") as dependent variable, the goal of this model is to establish the probability with which one observation can generate one value of the dependent variable rather than the other value. In this specific case, this model points out the probability for the student to spend more than five years (be "out of time") or not. Moreover, using the "margin" command in Stata, is possible to check if the probability to be "out of time" change when the number of gap years change as well. This is useful to understand if there are some differences between a student who had one gap year rather than more than one.

The equation of this regression is given by:

$$\begin{aligned}\Pr(Y = 1 \mid X) = F(&\beta_1 HSMark + \beta_2 y_{gap} + \beta_3 age \\ &+ \gamma_1 male + \beta_4 exam\_mark \\ &+ \beta_5 MotherEduc + \beta_6 FatherEduc \\ &+ \gamma_2 d\_geoplace1 + \gamma_3 d\_geoplace2 \\ &+ \gamma_4 d\_geoplace3 + \gamma_5 d\_geoplace4 \\ &+ \gamma_6 d\_geoplace5 + \beta_7 UniFeedback \\ &+ \beta_8 Old\_Uni + \beta_9 DegreePoint \\ &+ \beta_{10} Uni\_Work)\end{aligned}$$

Where, the function F is the cumulative logistic distribution.

Data

The data used in this work is from an Italian dataset which contain information about students. The dataset contains information on students who enrolled at the university from 2000 to 2010 and who completed their five years degree between 2010 and 2015. Moreover, the dataset contains information on the graduation date as well. The dataset contains only those faculties that require at least five years to be accomplished. Therefore, the undergraduate degree (three-year) and the postgraduate degree (two-year) are not includedin the dataset used for this project. The list of these faculties that directly offer the five-year postgraduate degree (instead of doing a three-year undergraduate plus a two-year postgraduate) is showed in the Table1[5] below.

[5] In theTable1 there are some codes in brackets. This code is used to distinguish the same faculty from the old system (LM) to the new system (S).

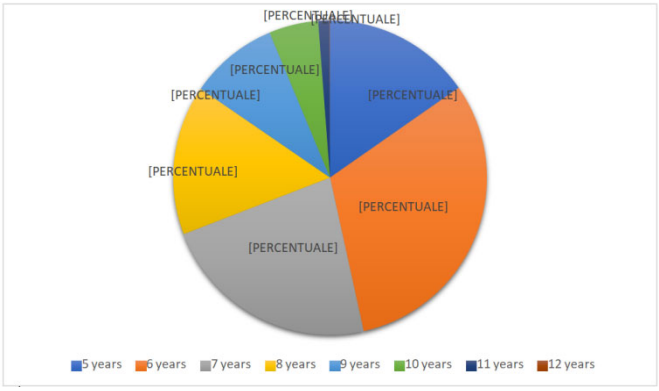
Table 1. List of faculties

Faculty	Freq.	Percent	Cum.
Biotechnology	2	0	0
Civil engineering	1	0	0
Informatics engineering	2	0	0.01
Economics science	2	0	0.01
Informatics science	1	0	0.01
Tourism	1	0	0.01
Civil engineering and architecture (4/S)	11875	12.91	12.92
Pharmacy (14/S)	17992	19.57	32.49
Medical veterinary (47/S)	3245	3.53	36.02
Dentistry (52/S)	2529	2.75	38.77
Science of architecture	1	0	38.77
Law	49008	53.3	92.07
Pharmacy (LM-13)	3307	3.6	95.67
Medical veterinary (LM-42)	955	1.04	96.71
Dentistry (LM-46)	8	0.01	96.72
Civil engineering and architecture (LM-4)	2923	3.18	99.9
Restoration of cultural heritage	96	0.1	100

The dataset contains a big number of variables: there are 250 variables. Because of this large number, can be useful to split all these variables in different groups to understand in a clearer way which of these can be useful for this project. There are variables about:

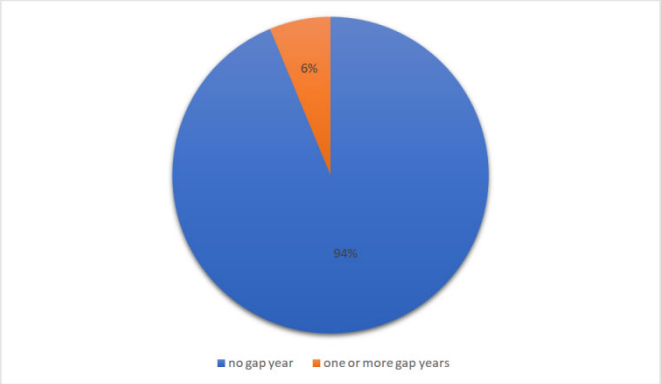
- Student personal information (age, sex, place of residence, nationality, etc...)
- Employment and education status of the parents (job and type of degree)
- High school graduation (mark, year, type of high school, etc…)
- University (faculty, place, time to degree, age of degree, average mark, mark of degree etc…)
- Abroad experience, e.g. ERASMUS
- Satisfaction of the students from the university
- If the student had a previous experience in another faculty
- Which aspect of a job are important for the students

For this work the more relevant variables are those that show information about students and their parents, to understand their background. Moreover, the variable about time to degree is the most important variable and is the one that is used as dependent variable; the name of this variable is “duration” that represent the time to degree, that is how long it takes the student to graduate. To show some descriptive statistics, we also created one other variable, that is the same value of “duration”, but rounded without decimal (“dur2”). In fact, if a student spends, for example, also one month more than six years this means that he is already enrolled in the seventh year. In the Graph1 below it is showed the time to degree of the students in the dataset, rounded as an integer, in the way that was pointed out above.

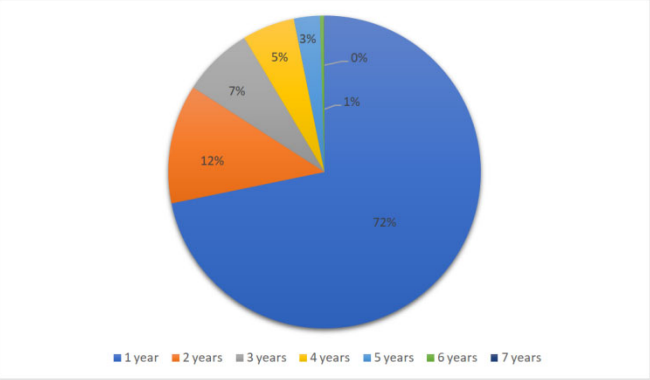


Graph 1. How many years it takes for a student to graduate

The dataset contains 76,336 students. As we can see from Graph 1, 31.25% of them need six years to reach their degree, meanwhile just the 15.31% can complete their studies in time. Moreover, just a very little part of students spent ten or more years to complete their path; in fact, the 93.70% of the observations are included between five to nine years. The rest of the observations can be considered as an unusual thing, because just a few of students spend more than nine years to reach the degree and for this reason STATA does not take in account these observations. Another important variable for this study is the gap year and moreover, how many years of gap the student decides to have. To evaluate this important topic, there are two useful variables in the dataset: “gapy” and “y_gap”. The former is a dummy variable that is equal to 1 if the student had a gap year and equal to 0 otherwise, the latter points out the number of years that the student waited before to enrol in the university. These variables are showed in the Graph2 and Graph 3 below.



Graph 2. Having a gap year or not



Graph 3. Number of years between the end of the high school and the enrolment in the university

The choice to have a gap year is not so usual; in fact, out of 76,336 just 4,774 students, that represent the 6.25% of the dataset, decide to not enrol immediately in the university. Analysing these students, it easy to understand that most of them (71.78%) had just one year of gap, that can be considered as the common choice for who decide to do not enrol immediately after the end of the high school. Instead, is less common for the students to spend more year: 12.32% of them spent two years and the percentage decreases as the number of years increase.

RESULTS

Using the statistic software STATA, it was possible to point out the results of the four tests described previously. The tests are showed in this order:

1. Linear regression: using the time spent by students to reach their degree ("duration") as dependent variable and some of the variables.
2. Kruskal-Wallis model: used to determine if two or more groups of the independent variable ("y_gap") are statistically significant different.
3. Chi-squared test: used to test for a relationship between two categorical variables.
4. Logistic regression: used to point out the probability for the student to spend more than five years or not.

Following this order, the first model was the Linear Regression, and in the Table2 below it is possible to see the STATA's output.

Table 2. Linear Regression output

Number of obs	4,137			
R-squared	0.7932			
duration	Coef.	Robust Std. Err.	t	P>|t|
HsMark	0.0047831	0.0010154	4.71	0
y_gap	0.9040614	0.01844	-49.03	0
male	0.0811238	0.0178211	-4.55	0
age	0.7893293	0.0105218	75.02	0
exam_mark	0.0543473	0.005527	-9.83	0
MotherEduc	0.048396	0.0153309	3.16	0.003
FatherEduc	0.0405431	0.0131739	3.08	0.004
geoplace1	0.2149198	0.0457177	-4.7	0
geoplace2	0.1867897	0.0367531	-5.08	0
geoplace3	0.0533824	0.0415657	-1.28	0.207
geoplace4	0.2368998	0.0268429	8.83	0
geoplace5	0	(omitted)		
UniFeedback	0.0110776	0.0071269	1.55	0.129
Old_Uni	0.1774746	0.0167062	10.62	0
DegreePoint	0.0002808	0.0001322	2.12	0.041
Uni_Work	0.0284326	0.151163	-1.88	0.068
_cons	-13.15797	0.3408404	-38.6	0

The R-squared is 0.7932 and this means that the model's inputs can explain almost the 80% of the observed variation, that is a quite good value. As is possible to see in Table 2, 16 variables were used to explain the model, just few of them are not statistically significant and the variable "d_zonageo5" is omitted to avoid the dummy variable trap. This trap is a scenario in which the independent variables are multicollinear; this means that two or more variables are highly correlated and so, one variable can be predicted from the others. To avoid this problem there are two possibilities: either drop one of the categorical variables or drop the intercept constant. In this case, the default solution adopted by STATA is the former one. In this regression the observations are clustered by the mark of degree. Taking in account the significant variables, theTable2 gives some important information. As the Table2 suggests, since the dummy variable for the gender ("male") has a negative coefficient, women are faster than men to accomplish their degree. Moreover, the variable that points out the work during the studies ("Work_tot") presents a negative coefficient, that can be seen as an indication that students who worked during their studies are slower in reaching their degree. Also, some variables that illustrate the background of the students are useful, in fact, the result shows that having parents with a higher level of education ("MotherEduc" and "FatherEduc") implies an increase of the time to degree ("duration"). A possible explanation for this result is suggested by Giannelli and Monfardini (2000 – 2003). According to their studies, the time spent to reach the degree is linked with the decision to live with parents, and if the father has a university degree the students are double more likely to live with their parents. The explanation given in their studies is that the parents can cover different cost, such as the tuition fees, and this allows students to reduce their effort and delaying the decision to form a family. Finally, analysing the most important variable for this study, looking the sign of the coefficient, it is possible to point out that when the number of gap years ("y_gap") increases the duration of the studies ("duration") decreases, since the coefficient is negative. This result can be seen as a confirmation of the main idea that is the starting point of this project. In fact, the idea is that students who start their path later, because they had one or more gap years, have an incentive to catch up their own coetaneous, who started before them and for this reason are closer to the degree and to get a job, as Lahey (2008) suggests. Going on with the next test, that as it was pointed out in the first paragraph of this chapter, is the Kruskal-Wallis test. This test shows if the population medians of the groups are statistically different among them respect the variable "duration", that is the continuous dependent variable for the time to degree.

The results are showed in Table 3.

Table 3. Kruskal-Wallis output

y_gap	Obs	Rank Sum	
1	3,419	8.84E+06	
2	588	1.36E+06	
3	349	695440.00	
4	259	348347.50	
5	128	99659.00	
6	21	10529.00	
7	2	180.00	
chi-squared =		469.99	with 6 d.f.
probability =		0.0001	

In the last two rows of the output are highlighted the final result. Before to explain these results, it is important to write down the null hypothesis (H_0) and the alternative hypothesis (H_1):

- H_0:Population medians are equal
- H_1:Population medians are not all equal

The first row shows the chi-squared value, that is 469.99, and the degrees of freedom, that are six in this specific case. Moreover, the last row reports the p-value of the test; in this test the statistical significance is 0.0001, that since is below the value equal to 0.05, indicates that it is possible to reject the null hypothesis, according to which the population medians are equal, and so there is a statistically significant difference among the groups. So, this second test is a confirmation of what the first one already pointed out. The latter, as it was already described above, shows that when the number of gap years ("y_gap") increases the time to degree ("duration") decreases. Moreover, the former reports that having a different number of gap years ("y_gap") imply spend a different time to reach the degree ("duration"). These two tests are saying almost the same thing, because both of them describe a correlation between the two main variables. The difference is that meanwhile the first test shows how one variable affect the other, the second one just points out that there is an effect between them. After the first two tests, to analyse in a different way this topic is useful to use the other two tests with the dummy variable "outtime" as the dependent variable. The first one is the Chi-squared test for independence. The Table4 shows the STATA's output of the test.

Table 4. Chi-squared test output

	Gapy			
Out time	0	1	Total	
0	10,775	916	11,691	
	10,959.10	731.9	11,691	
1	60,584	3,850	64,434	
	60,399.90	4,034.10	64,434	
Total	71,359	4,766	76,125	
	71,359	4,766	76,125	
Pearson chi2(1) =		58.3322	Pr =	0.000

This is a two-way tabulation, of two dummy variables. The variable in the colunm is 0 for the students who spent at maximum five years to graduate and 1 for the students who spent more than five years ("outtime"). Meanwhile, the variable in row is 0 if the student does not have the gap year and 1 otherwise ("gapy"). As reported in the description in the other paragraph, this test will compute the expected theoretical frequency of the two variables, supposing that are independent. This frequency corresponds to the expected frequency, that are the number below in each cell. Instead, the other number in each cell, corresponds to the frequency observed in the data. After this analysis, it is important to state the null hypothesis (H_0) and the alternative hypothesis (H_1). Then, in this case:

- H_0: spending more than five years and having a gap year are independent
- H_1: spending more than five years and having a gap year are not independent

The last row in Table 4 shows the result. First of all, in that line is pointed out inside the brackets the degree of freedom, that is one in this case; then, there is the value of the test, 58.3322; finally, there is the p-value, that is the easier way to understand the result of the test. Since the p-value is less than 0.05, that is the significance level, it is not possible to accept the null hypothesis. So, it is reasonable to suppose that there is a relationship between being late and having a gap year. The last test showed that there is a relationship between the main variables of this study, but it was not enough to explain in which way these two variables affect each other, in fact is not possible to say if the relationship between this two variable is neither directly proportional nor inversely proportional. For this reason, is important to carry on the Logistic Regression model.

Table 5. Logistic Regression output

Number of obs	4,136			
Out time	Coef.	Std. Err.	z	P>|z|
y_gap				
2	-2.64636	0.200158	-13.22	0
3	-5.23827	0.277422	-18.88	0
4	-7.83222	0.352004	-22.25	0
5	-10.5095	0.440647	-23.85	0
6	0	(empty)		
Hsmark	0.0121	0.005276	2.29	0.022
foreignStud	0.283698	0.148715	1.91	0.056
male	-0.24107	0.115936	-2.08	0.038
age	2.186397	0.089702	24.37	0
exam_mark	-0.26577	0.035602	-7.46	0
TitMother	0.240913	0.078708	3.06	0.002
TitFather	-0.04555	0.076514	-0.6	0.552
geoplace1	-0.53389	0.237698	-2.25	0.025
geoplace2	-0.25245	0.197766	-1.28	0.202
geoplace3	0.167823	0.198106	0.85	0.397
geoplace4	0.974709	0.190589	5.12	0
geoplace5	0	(omitted)		
Old_Uni	0.378119	0.108613	3.48	0
DegreePoint	0.001936	0.000697	2.78	0.005
Uni_Work	-0.28993	0.11485	-2.52	0.012
_cons	-49.9365	2.567826	-19.45	0

In this regression, as it is pointed out in the Table5, the variables used are almost the same of the one used in the first linear regression presented in this study. Most of the variables are statistically significant and again the variable "d_zonageo5" is omitted to avoid the dummy variable trap. As was pointed out before, this trap is a scenario in which two or more variables are highly correlated and so, one variable can be predicted from the others. To avoid this problem STATA drops one of the categorical variables. Since the dependent variable is dummy, in this model is pointed out how the independent variables influence the possibility that the student spent more than five years to reach the degree ("outtime"). Focusing on the main variable that shows how many years the student waited before to join the university ("y_gap"), this result shows that the sign is negative. This mean that having two, three, four or five gap years decreases the possibilities to take more years to accomplish the degree. Unfortunately, due to the few observations, STATA omitted the students with more than five years of gap. Omitting outliers is something that can happen during a study for different motivations; it is important to understand why there are some outliers, because sometimes is not possible to just omit these observations. Vice versa, if the outliers are due to measurement error or unusual conditions, the researcher can decide to leave them out. In this project these outliers are represented by just a few students. Probably, having more than five gap years it is an unusual condition and can be due by a mistake in filling the data as well. For this reason, it is important to minimize and prevent possible entry errors removing these outliers, as suggested by Kwak and Kim (2017). With this first output it is possible to make a statement about the sign of the coefficient and it is possible to analyse the effect of having or not the gap year. To analyse in a deeper way the effect on the dependent variable it is used the command "margin"; in this way it is possible to decompose the variable to understand not just the effect of having or not the gap year, but the effect per each number of gap years as well. The output of this command is showed in Table 6.

Table 6. Prediction output

Expression:	Pr(outtime), predict()			
Number of obs	4136			
	Margin	Delta-method Std. Err.	z	P>|z|
y_gap				
1	0.880207	0.003852	228.5	0
2	0.68594	0.013711	50.03	0
3	0.456631	0.016079	28.4	0
4	0.246193	0.013251	18.58	0
5	0.079093	0.010783	7.34	0

This last output is in addition to the logistic regression and it is helpful to understand in which way the number of gap years affects the possibility to spend more than five years to take the degree. Also, in this output the students with six or more gap years are not taken in account since this output depends on the previous one. The "margin" column shows the probability for each group of the variable to be "out of time". So, the probability to be out of time if the student had one gap year is higher than the same probability for a student who had two gap years, and so on. This means that when the number of gap years increases the probability to be out of time decreases. To sum up, this last test confirms what was already shown with the previous tests: when the number of gap years increases the time to degree decreases and having different number of gap years imply different effect on the duration of studies. Moreover, the result of the fourth test seem to confirm what

was pointed out by the third one, due to the fact that also in this last test there is a relationship between gap years and being late or not. Since both tests point out the same result, it is possible to be more confident on the conclusion. In conclusion, all the tests pointed out in this project were useful to confirm the main idea behind this study: students who had one or more gap years, are more motivated to try to reach their degree as fast as they can, as Martin (2010) suggests. In fact, Martin (2010) shows in his study that the participation to the gap year can predicts a higher academic motivation. Probably, this event can be due to the fact that students know that they started late their path and so they try to be faster to make up for time that they spent in other ways, instead of enrol immediately in the university. Moreover, the thought that other students of the same age, who did not have the gap year, are in advantage, can be another motivation to try to be faster, mainly because, as it was pointed out by Lahey (2008), after the university, a younger worker who is looking for a job is 40% more likely to have an interview than an older one.

CONCLUSION

During the last few years, the number of students who had an extension of the time to reach their degree in Italy has increased. This has attracted a lot of attention from policy makers and practitioners because students represent the future human capital available for a country and can be an important proxy to understand how the country can growth in the future.
So, it is necessary to help the students to develop their skill and to reach their degree without an extension of the time. The main reason because this is a problem for the students themselves, is because a student who took his/her degree spending more time will be an older worker with less potential growth. But, before to help them in their studies, it is important to understand what they really need. The main reason to develop a system that can be helpful for the students is due to the fact that these people represent the future human capital in the job market, and an improvement in human capital become an improvement for the country itself. The second phenomenon taken in account in this project is the gap year. This time, during the gap year, can be used by the students for many reasons and nowadays the number of students having a gap year is increasing. Some of the reasons to have a gap year are starting a part-time job, having some rest, doing some volunteering. As it was pointed out in the second chapter of this study, there is a correlation between the gap year and student performance, and the satisfaction of the students is affected as well. For these reasons, since there is the willingness to improve the human capital, this phenomenon can be taken in account to figure out some possible development. The goal of this study is to link the "time to degree", that is the time spent by the students to reach the degree, to the gap year and to understand if there is a relationship between these two phenomena and how this relationship works. To do so, in the third chapter are highlighted the results of the methods used to analyse this relationship. These tests show that there is a dependence between having the gap year and spend less than five years and, moreover, the number of gap year is linked with the duration of the studies as well. According to these results it is possible to say that having a gap year between high school and university can be helpful for the students to reach a greater awareness of their self and to really understand what they what to reach in their life and, as a consequence, to be more efficient during their studies. In this study the time spent from the students to reach their degree is used as a proxy of the efficiency. In fact, a student who take the degree when is younger, is a younger worker as well and this can be seen as a worker with a high margin for potential growth. Moreover, as already pointed out in this study, take the degree in a younger age it means an increase in the possibility to find a job for the students. To sum up, this study is focused on the possibility of using the gap year to generate an improvement on the students' path, to improve the human capital consequently.

Finally, it is also important to point out that what is described until now is something that does not need a monetary investment by the government or the university, because the students decreased their "time to degree" just because they had a gap year. So, the reason that bring this decrease of the time spent by the students to reach the degree it was a personal development of the students and not something tangible like a greater number of libraries or a scholarship. This fact leads to take in account the behaviour economics that, as opposed to the classic economics, is based to some human factors, as the will or the instinct, that lead to analyse this phenomena from a psychologist point of view. As a consequence of this analysis, there are some potential avenues for future research. In fact, a possible development for this study can be to analyse if the link between the gap year and the time to degree can be extended to the time to find a job as well. Moreover, instead of use the performance as a student can be interesting to use the performance as a worker. In this way it can be possible to understand if the gap year has an effect not just during the university's life but during the entire working life of the person. To sum up, it is important to analyse in a deeper way this topic, in order to find other solutions, not just in the economic field, that can bring an improvement of the human capital and for the students' life as well.

REFERENCES

Aina, C. and Pastore, F. (2012), "Delayed Graduation and Overeducation: A Test of the Human Capital Model versus the Screening Hypothesis", IZA Discussion Papers, n. 6413

Aina, Carmen, and Giorgia Casalone. "Does time-to-degree matter? The effect of delayed graduation on employment and wages." Vol. 38. AlmaLaurea working papers, (2011).

Becker, Gary S. "Investment in human capital: A theoretical analysis." *Journal of political economy* 70.5, Part 2 (1962): 9-49.

Becker, Sascha O. "Why don't Italians finish university?." Explaining enrolment behavior in Italy and Germany. Available online at http://www. sobecker. de/dropout. pdf (2001).

Becker, S. (2006), "Introducing Time-to-Educate in a Job Search Model", *Bulletin of Economic Research,* 58(1): 61-72

Bound, John, Michael F. Lovenheim, and Sarah Turner. "Increasing time to baccalaureate degree in the United States." *Education Finance and Policy*, 7.4 (2012): 375-424.

Bound, J. e Turner, S. (2007), "Cohort Crowding: How Resources Affect Collegiate Attainment", *Journal of Public Economics,* 91, 877-99

Brodaty, T., Gary-Bobo, R.J. e Prieto, A. (2008), "Does Speed Signal Ability? The Impact of Grade Repetitions on Employment and Wages", Center for Economic Policy Research (CEPR), Discussion Papers, n. 6832.

Chan, Yvonne, and Roy P. Walmsley. "Learning and understanding the Kruskal-Wallis one-way analysis-of-variance-by-ranks test for differences among three or more independent groups." *Physical therapy* 77.12 (1997): 1755-1761.

Coetzee, Melinde, and Suzanne Bester. "The possible value of a gap year: A case study." *South African Journal of Higher Education* 23.3 (2009): 608-623.

Giannelli, G., and C. Monfardini. "Young Adults, Family and the Labour Market: An Empirical Analysis of Human Capital Investment Decisions." *International Journal of Manpower* 21.3-4 (2000): 227-45.

Hamir, Heather Barclay. Go abroad and graduate on-time: Study abroad participation, degree completion, and time-to-degree. The University of Nebraska-Lincoln, (2011).

Krause, Kerri-Lee, *et al.* "The first year experience in Australian universities: Findings from a decade of national studies." (2005): 31-44.

Kurzweil, Martin, and D. Derek Wu. "Building a pathway to student success at Georgia State University." (2015).

Kwak, Sang Kyu, and Jong Hae Kim. "Statistical data preparation: management of missing values and outliers." *Korean journal of anesthesiology,* 70.4 (2017): 407.

Lahey, Joanna N. "Age, women, and hiring an experimental study." *Journal of Human resources,* 43.1 (2008): 30-56.

Martin, Andrew J. "Should students have a gap year? Motivation and performance factors relevant to time out after completing school." *Journal of Educational Psychology*, 102.3 (2010): 561.

McHugh, Mary L. "The chi-square test of independence." Biochemia medica: *Biochemia medica* 23.2 (2013): 143-149.

Mortada, L., J. Bolbol, and S. Kadry. "Factors Affecting Students' Performance a Case of Private Colleges in Lebanon." *J Math Stat Anal* 1 (2018): 105.

Parker, Philip D., *et al.* "I wish I had (not) taken a gap-year? The psychological and attainment outcomes of different post-school pathways." *Developmental Psychology* 51.3 (2015): 323.

Pastore, Francesco. "School-to-work Transitions in Italy. A steeplechase with no Winner?." XXIV AIEL Conference (2009).

Rose Birch, E. L. I. S. A., and Paul W. Miller. "The characteristics of 'Gap-Year'students and their tertiary academic outcomes." *Economic Record* 83.262 (2007): 329-344.

Savickas, Mark L. "Annual review: Practice and research in career counseling and development, 1988." *The Career Development Quarterly* 38.2 (1989): 100-134.

Savickas, Mark L. "The theory and practice of career construction." Career development and counseling: Putting theory and research to work 1 (2005): 42-70.

Wright, Bryan E. "Use of chi-square tests to analyze scat-derived diet composition data." *Marine Mammal Science* 26.2 (2010): 395-401.

****