LEARNING ANALYTICS TO EVALUATE THE EFFECTIVENESS OF HIGHER EDUCATION STUDENT FAILURE PREVENTION

Francesco Floris[1], Marina Marchisio[1], Matteo Sacchet[1] and Sergio Rabellino[2]
[1]*Department of Molecular Biotechnology and Health Sciences, University of Torino*
Via Nizza 52, 10126 Torino, Italy
[2]*Department of Computer Science, University of Torino*
Via Pessinetto 12, 10149 Torino, Italy

ABSTRACT

Open Online Courses can serve different purposes: in the case of Orient@mente at the University of Torino, they aim at facilitating the transition from secondary to tertiary education with automatic evaluation tests that students can try in order to understand their capabilities in - and their attitude towards - certain disciplines, and with remedial courses to fill the gaps in their knowledge. The university strategy of Orient@mente first started in 2014 and, after six years of deployment of the online platform, it has collected many data from students that were interested in starting a university course. The natural comparison and correlation analysis juxtapose the academic results of students who practice self-assessment in Orient@mente with the other university students. The measurable that we considered in this comparison is the average number of ECTS acquired during the first year, since this is a number that is considered for the university evaluation system - stakeholders in particular check the number of students who obtain more than 40 first-year ECTS. According to specific rules, we put together dataset from different origins, the platform logs, and users with the university record system. The results of this analysis, presented in this work, confirm the positive impact of Orient@mente on students, with statistical significance.

KEYWORDS

Data Analytics, Learning Analytics, MOOCs, Secondary to Tertiary Transition, University Guidance, University Orientation.

1. INTRODUCTION

Massive Open Online Courses (MOOCs) are a worldwide instrument to face diverse aspects in education: open education, open access, interactivity, and digital competencies. The education topics need many pedagogical studies on user interaction with online systems. MOOCs can pursue several goals, which are not necessarily strictly related to disciplinary contents. By using an automatic evaluation system, it is possible to enhance students' ability to self-evaluate and manage their self-regulated learning. By using computing environments, it is possible to enhance problem solving skills. By using web conference tools, it is possible to enhance collaborative competencies across different regions and countries.

In the 2010 report of the European Commission "Europe 2020" (European Commission, 2010) one of the objectives to be achieved for smart, sustainable and inclusive growth is the reduction of the abandonment rate (bringing it below 10%) and the increase in the percentage of young graduates up to 40%.

This objective is to lower the drop-out rate by five percentage points (in 2010 the European average was around 15%) in 10 years and increase the share of the population aged between 20 and 34 who successfully completes high school by approximately 10 percentage points.

The achievement of this goal is closely connected with the problems that students encounter in the transition phase from secondary to tertiary education, what is known as the Secondary to Tertiary Transition (STT). It is a decisive change in students' perspectives, since they need to be aware of the choice that will affect their future and be prepared to face the difference between expectations and reality. It is not surprising that first-year students are marked as unprepared, unable, or unwilling to cope with challenges of university-level

education. From the student side, the STT experience looks differently. Students can face unexpected failures of the learning strategies they developed by previous education. The education system is responsible for making this challenging experience as positive and smooth as possible. Technologies and virtual environments are key components of the support to students, because multimedia resources give students an overview of what they will face at university in advance, communication technologies allow university students to interact with professors, online tests containing feedback allow students to check whether their preparation is adequate to attend first year courses, and interactive learning materials can help them fill their gaps.

Through various initiatives, universities have tried to adapt to the situation, developing different guidance activities, covering the various study programs. The main objective of university orientation is the reduction of the dropout rate, which directly impacts on the evaluation of the university itself, both in quantitative and qualitative terms. Another objective of universities is the balance between the education of new professionals and the demand of the job market. Because of this, many degree programs require students to take an admission test in order to assess the minimum requirements and provide a restriction on the numbers of new students, according to the score they get from questions about different topics related to the degree program. Orientation days, job fairs and open days are the most common initiatives promoted by universities, education offices and secondary schools. In addition to these orientation initiatives, new technologies offer a solution for the mass dissemination of information and assessment to facilitate the orientation of students, mediated by Learning Management Systems (LMS), which contains multimedia resources, interactive components, and online tests.

University of Torino proposes several strategies and actions to facilitate the transition from secondary school to university through the use new technologies. The common model consists in an open online platform, integrated with an Advanced Computing Environment (ACE), an Automatic Assessment System (AAS) and a Web Conference tool, that delivers self-paced Open Online Courses (OOC), even if not Massive since the main focus of these strategies is on the student. Moreover, there is no restriction on when to access courses: students can start them anytime. Considering the users of the various actions, more than 80,000 users are involved. Some of them can be duplicate across the platforms, but around 60,000 distinct students subscribed, from the most populated platform. On average, every month there are around 2,000 unique logins which correspond to more than 60,000 sign-ins per month.

Thus, the online environments collect a lot of data, which are very useful to analyze and evaluate the initiatives and improve and enhance their effect. The main criterion for assessing the effectiveness of these actions is the number of ECTS obtained by students in their first year of academic career, in relation to the online activity. In the present work, we will discuss the analysis of data collected from different sources of the action Orient@mente, with a methodology that can be applied to various strategies adopted at university level.

2. STATE OF THE ART

Every student failure is a failure of the whole education system, at any level this may occur. Thus, preventing failures and enhancing success are key objectives of all the institutions. The various strategies that can be adopted address evaluation, grouping, motivation, practices, responsibility, and relationships. The possible interventions can be divided into two main branches (Krull and Duart, 2018): academic support, which deals with the development of cognitive and learning skills, and non-academic support, which deals with emotional and organizational aspects such as counselling and guidance services. Both aspects have a positive impact on the students' experience.

The decision on which changes or strategies are good for preventing failure and enhancing success is driven by descriptive analytics, but in a fast-changing world predictive and prescriptive models. The ability to detect possible failures in advance allows preventive interventions. There are many experiences of prediction algorithms: the latest trends use data mining techniques (Marquez-Vera et al., 2013) and Learning Analytics (LA), as one can see from the trend topics in related conferences (Ochoa and Merceron, 2018). These approaches also make use of machine learning algorithms, which are generally used as black boxes which provide an output without any information on the decision process. More research needs to be devoted to the ethical aspects, since they need to be thoroughly examined in terms of fairness (Riazy et al., 2020), given that predictions of a student's performance have an impact on the student's success. Researchers developed methods and tutorials to detect discrimination (Hajian et al., 2016), which is endemic even when there is no discrimination intention in the developer of the algorithm.

These strategies for preventing failure and improving success find a strong point of support in the decision to use numerous guidance strategies by University of Torino through Learning Management Systems, even if this is not the standard for Italian universities. In fact, from the OECD (OECD/European Union, 2020) emerges the present perspective about Higher Education (HE) in Italy, and the various priorities of HE institutions. From the OECD HEI Leader Survey, to the question "Taking into account the HEInnovate dimensions/components listed below, please indicate the three that are most prominent in your strategy", none of the survey respondents considers the enhancement of digital transformation as one of the top three priorities for the HEI. This is made worse by the allocation of most recent investments in digital technologies and systems performed by Italian HEIs, the lowest amount of which is devoted to developing digital skills of professors and researchers, who are directly responsible for the education of students. It is worth mentioning that the survey was not administrated to online universities. In this context, the following orientation actions of the University of Torino take place.

Orient@mente (https://orientamente.unito.it) consists of OOCs for the interactive exploration of courses, for the preparation for tests for the recovery of gaps in many courses, from natural sciences to humanities. It serves as an effective dematerialized orientation for secondary school students who will enroll to university. After a proper training and support in the use of the technologies, university professors select and arrange the materials in the courses, as a guarantee of quality (Barana et al., 2017).

Start@unito (https://start.unito.it) promotes and facilitates the transition from secondary school to the university system through the creation and dissemination of a series of OOCs, which are full university modules in many different first year disciplines. As a second effect, it helps students make a responsible choice about academic studies by offering interactive paths for university guidance and preparation for admission tests. Moreover, Start@unito addresses two non-traditional types of learners, namely university professors and high school teachers. These non-standard learner groups are considerably smaller than the traditional student population, but they have a large impact in the education sector (Marchisio et al., 2019).

Foundation Programme (https://foundationprogramme.unito.it) facilitates the integration of pre-university students with less than 12 years of education. The additional semester or year allows the achievement of the corresponding ECTS, according to international students' needs. Since it is delivered completely online, students can attend the program from their home country, without the need of a permission or visa and without travelling. Only the final exam that allows to certify the ECTS has to be taken at the university, right before enrollment (Floris et al., 2019).

3. RESEARCH QUESTIONS AND METHODS

The goal of our analysis is to understand the impact of the Orient@mente platform on the career of students. Summing up previous research, with Orient@mente (Barana et al., 2017) the strategies to prevent student failure concern both the learning experience, with automatic assessment and courses, and the organizational one, with information on courses and interviews to older students or professors.

RQ1. How to evaluate the impact of the orientation action of Orient@mente on students' careers?

RQ2. How to use Learning Analytics (LA) in order to obtain a suitable evaluation of the impact of Orient@mente action?

All the university actions take place in a fully online environment, addressing the increasing need of digitalization of HE institutions. Data about the usage and the advantages of the online platform have already been studied in previous works: the originality of this research lies in the direct measurement of the correlation between academic results of students who used Orient@mente, tested against students who did not. It is important to measure the possible improvement of students who used the platform in admission tests and exams. The analysis considers data about the careers of students enrolled at the University of Torino from the degree programs that have a course in the Orient@mente platform, and students' data arising from the interaction with the platform.

Hypothesis (H0): is there a significant difference in students' academic performance in relation to their activity on the Orient@mente platform?

Data about academic performance consist in the weighted average grade and the number of University Credits (ECTS) acquired by students at the end of each exam session and at the end of the first year of university. This measurables are the most relevant, since the University evaluation system makes use of similar

indicators: in particular, it evaluates the number of students that gain more than 40 ECTS, which is in fact an excellent aggregator for the dimensions of students who continue their studies and productivity. The analysis considers the previous education of students, such as the type of school attended, the city, the region, and the graduation mark. Data about the platform are divided into two main areas: the first one comes from students interaction with the LMS, while the other data source is the database of the AAS, which contains all the necessary information about tests, such as grades, number of attempts, time, and other data.

RQ2 will be investigated more qualitatively, analyzing which aspects of the adopted analysis methods provided useful insights and which one were avoidable, according to precise indicators (Marchisio et al., 2019; Scheffel et al., 2014).

4. ANALYSIS PROCESS

Data analysis was divided into different phases.

- Phase 1 (P1). The first action on the various datasets was merging them and creating a unique large database. The connecting point between the different databases is given by the email field, which may not be unique. The data about the students' career at the university contain two email addresses, the private and the institutional ones.
- Phase 2 (P2). The second action on the merged dataset was cleaning and transforming. This is an important phase which gives an overview of the whole population and prepares the data for the further phases (grouping and statistical analysis).
- Phase 3 (P3). Statistical hypothesis testing, with t-test and chi-square test, to verify the validity of the hypothesis (H0).
- Phase 4 (P4). The clean dataset can be grouped according to some variables expressing the use of the platform, like the number of logs and the number of tests, in correlation with the student's achievement at university level.

Two software were used for this analysis: "KNIME Analytics Platform 4.1.1" and "SAS 9.4". "KNIME", used in (P1), allows, with the node system, easy access to different data sources (career data, platform database), and the use of SQL language to access the necessary information and prepare the dataset for next phases. The merged dataset was imported to "SAS" for (P2) and (P3), for cleaning, transformation of database and statistical analysis. Last phase (P4) was performed on "KNIME" again. In this paper we will focus on the first three phases.

5. RESULTS

The analyzed sample consists of students who enrolled at the University of Torino from 2010 to the academic year 2017/2018.

The size of the dataset is 29256 observations, divided into two subsamples: the first one (SS1) composed of students that are Orient@mente users, too, containing 22.38% of the global sample, while the second one (SS2) consists of students who are not Orient@mente users. There are no significant differences in the proportionality of genders in both subsamples. Table 1 describes the percentage of users over the global number of observations, divided into (SS1) and (SS2) and grouped by the leaving certificate grade, for which the maximum grade is 100.

From Table 1 we can see that the distribution of students across the various ranges of grades is quite uniform, with percentages around 20% and 30%. Even inside (SS1) and (SS2) the same global proportion is respected. A similar procedure was carried out about gender, with no statistically significant difference in the proportion of males and females (Table 2).

Data do not show the same proportionality about the educational origin (Table 3), because it emerges that students from the "Liceo Scientifico" high school are less interested in attending Orient@mente, and this can be explained with the fact that most of the MOOCs available are about science, which is a pivotal discipline in that kind of school.

Table 1. Percentages and numbers of users according to the subsample and grouped by the leaving certificate grade

Group	**SS1 (freq, pct, pct col)**	**SS2 (freq, pct, pct col)**	**Row total**
R1: First range of students with leaving certificate grade between 60 and 69	1366 students	5023 students	6389 students
Percentage	4.67%	17.17%	21.84%
Column percentage	20.86%	22.12%	
R2: Second range of students with leaving certificate grade between 70 and 79	2029 students	6851 students	8880 students
Percentage	6.94%	23.42%	30.35%
Column percentage	30.99%	30.17%	
R3: Third range of students with leaving certificate grade between 80 and 89	1747 students	6081 students	7828 students
Percentage	5.97%	20.79%	26.76%
Column percentage	26.68%	26.78%	
R4: Fourth range of students with leaving certificate grade between 90 and 100	1406 students	4753 students	6159 students
Percentage	4.81%	16.25%	21.05%
Column percentage	21.47%	20.93%	
Column total	**6548 students**	**22708 students**	**29256 students**
	22.38%	**77.62%**	**100%**
	100%	**100%**	

Table 2. Percentages and numbers of users according to the subsample and grouped by gender

Gender	**SS1**	**SS2**	**Row total**
Male	3246 students	11160 students	14406 students
Row percentage	22.53%	77.47%	100%
Column percentage	49.58%	49.15%	49.24%
Female	3302 students	11548 students	14850 students
Row percentage	22.24%	77.76%	100%
Column percentage	50.42%	50.85%	50.76%
Column total	**6548 students**	**22708 students**	**29256 students**
Row percentage	**22.38%**	**77.62%**	**100%**

Table 3. Percentages and numbers of users according to the subsample and grouped by educational origin

Educational origin	**SS1**	**SS2**	**Row total**
International	85	124	209
Percentage	0.29%	0.42%	
Row percentage	40.67%	59.33%	
Column percentage	1.30%	0.55%	0.71%
Professional School	437	1314	1751
Percentage	1.49%	4.49%	
Row percentage	24.96	75.04%	
Column percentage	6.67%	5.79%	5.99%
Istituto Tecnico	1638	4432	6070
Percentage	5.60%	15.15%	
Row percentage	26.99%	73.01%	
Column percentage	25.02%	19.52%	20.75%
Liceo	3977	14922	18899
Percentage	13.59%	51.00%	
Row percentage	21.04%	78.96%	
Column percentage	60.74%	65.71%	64.60%
Unknow	411	1916	2327
Percentage	1.40%	6.55%	
Row percentage	17.66%	82.34%	
Column percentage	6.28%	8.44%	7.95%
Column total	**6548 students**	**22708 students**	**29256 students**
Row percentage	**22.38%**	**77.62%**	**100%**

First access of 68.10% of (SS1) students was prior to enrollment at the university. 79% of (SS1) students did not use university institutional mail to access Orient@mente, even when the first access to the platform occurred after enrollment (16% of SS1). Thus, it is possible that in (P1) some (SS1) students do not correctly appear in the merged dataset, because it is possible that they used two different personal emails, one to subscribe to Orient@mente and one to enroll at the University of Torino. In cases like this one, the platform user cannot be traced back to the university student and will thus appear as a student in (SS2) in the dataset.

5.1 Hypothesis Tests

We performed a hypothesis test to check the equality between the average number of ECTS (which we will indicate with the Greek letter μ) achieved by the students at the end of the first academic year in the subsamples (SS1) and (SS2).

$$\begin{cases} H_0 : \mu_{SS1} = \mu_{SS2} \\ H_A : \mu_{SS1} > \mu_{SS2} \end{cases}$$

The p-value for the test much less than 0.0001, thus rejecting the hypothesis of equality, it is possible to state that there is statistical difference between the average number of ECTS achieved depending on the access to the Orient@mente platform.

The same Hypothesis test checked the equality of the weighted average grade of students between the subsamples (SS1) and (SS2). The p-value for the test is almost null, thus rejecting the hypothesis of equality. Thus, even for this second measurable there is statistically significant difference.

The behavior of the same measurables can be different according to the various groups of students. The same analysis was performed with a restriction of the sample to the four ranges (R1), (R2), (R3), (R4). There is statistically significant difference between the average number of ECTS of (SS1-R1), (SS2-R1) and of (SS1-R4), (SS2-R4) with p-value less than 0.005, while for (R2) and (R3) we cannot reject the hypothesis of equality. In particular, the case of (R3) is very significant, since difference between the two averages (SS1-R3) and (SS2-R3) is below 0.05

Thus, from the sample we analyzed, we can infer that Orient@mente action is more effective when it involves users with the highest and users with the lowest leaving certificate grades, two completely diverse kinds of students. Looking more closely at the behavior of students in these two ranges, 35% of (R1) students log in after enrollment, against 25% of (R4) students, showing that (R1) students use the platform to fill the gaps that emerge during the first year of studies. There is no particular evidence regarding the various platform activities carried out by these students, whose study regards both the test area and the realignment courses in the same way.

5.2 Focusing on Two Courses

There could be a strong difference between the various courses. Two cases were of interests for their different behavior. Case 1 (C1) concerns students in Mathematics (1271 students), for whom the average number of ECTS achieved by SS2 students is even 3.5 points higher (p-value < 0.02). This result is not surprising since a student who enrolls in the course Mathematics is usually someone who is aware of his own interests. Moreover, it is a student with no need to attend courses to remediate gaps in the discipline.

On the other hand, Case 2 (C2) concerns students in Biological science (2041 students), for whom the average number of ECTS earned by (SS1) students is 4 points higher compared to (SS2) students (p-value < 0.0001). This result is very important because 4 ECTS more mean an extra exam during first year. This is particularly relevant for courses where mathematics, physics, chemistry are not the main topics, but basic disciplines anyway. They often represent first-year stumbling blocks because they are the first modules that students encounter, but of less specific interest.

5.3 Features of the Adopted LA

There are different indicators that can be used to evaluate LA. We will briefly describe and summarize Structure, Background, Statistic methods, Population, Procedure, Availability, Outcomes, Intervention.

From the point of view of Structure, we have provided a full description of the kind of analysis that we adopted, pointing out the various phases which delineate the Procedure. The Background is provided by the Orient@mente action, which involved a large Population of prospective students that became university students afterwards, in comparison with students that did not use the orientation tool. The Statistic methods are given by the descriptive statistics of the sample and the hypothesis test to evaluate the equality between the average number of ECTS in the subgroups. The Availability of this analysis is a little bit restrictive, because it uses data from different sources that may require official procedure to access. Moreover, new data arises after years of activity. The Outcomes provide clear information: acceptance or rejection of the hypothesis, with statistical significance. These provides information to the institution in order to take data driven decision.

In this particular setting, given the results, the analysis and the evaluation of the impact of Orient@mente, the University of Torino took immediately action and has decided to invest in this initiative: new Orient@mente courses and new test areas are currently under development for more than 85% of the university degree programs and will soon become available to prospective students.

6. CONCLUSION

From the analysis, it emerges that the research hypothesis has statistical significance in the case of both the weighted average grade and the average number of first-year acquired ECTS, thus proving that there is a positive impact on students' outcome thanks to the OOC orientation actions, especially in the case of Orient@mente. In this work we provided a method which can be useful to analyze other systems and platforms in order to assess effectiveness (RQ1). Apart from the different sources of data, the method could be refined in order to be automatically executed and return real time results.

These results guide future university orientation actions (RQ2). In fact, the Orient@mente platform is going to be further enlarged with new courses, which are currently under development and will be available to students who are thinking about or going to enroll to one of the numerous degrees at the University of Torino. As described in subsection 5.3, the institution invested in new courses and new test areas, that are currently under development.

There are many more parameters that can guide the evaluation process over the guidance service and which can be included in close future research:

- Students' information in their personal profile to provide a personalized orientation experience, suited to their needs, to offer them quickly the content they need most.
- Design personalized learning paths using Machine Learning techniques to differentiate students' learning paths, adapting them to their attitudes and their learning speed.
- Cluster analysis to identify "behavior" patterns within the platform and study their effects. In this way, once recognized non-productive behaviors by new users, it is possible to intervene to suggest a more effective learning path.
- Learning Analytics techniques to connect the results of tests carried out with the automatic assessment system to improve the implementation of automatic formative assessment strategies and provide immediate and personalized feedback to students.
- Analysis of other variables, such as the region of the school, the type of school attended, the chosen program.

ACKNOWLEDGEMENT

The authors would like to thank the bank foundation Compagnia di San Paolo, which financially supports many of the initiatives at the University of Torino, in particular the project OPERA Open Program for Educational Resources and Activities, in which this research took place. The authors would like to thank Dr. Massimo Bruno, Director of the Didactics office at the University of Torino and his staff for the support they gave to the action Orient@mente.

REFERENCES

Barana, A., Bogino, A., Fioravera, M., Marchisio, M., & Rabellino, S. (2017). Open Platform of Self-paced MOOCs for the continual improvement of Academic Guidance and Knowledge Strengthening in Tertiary Education. Journal of E-Learning and Knowledge Society, 13(3). https://doi.org/10.20368/1971-8829/1383

Barana, A., Conte, A., Fissore, C., Marchisio, M., & Rabellino, S. (2019). Learning Analytics to improve Formative Assessment strategies. Journal of E-Learning and Knowledge Society, 15(3), 75-88. https://doi.org/10.20368/1971-8829/1135057

Floris, F., Marchisio, M., Marello, C. & Operti, L. (2019). Bridge the gap between high school systems with less than twelve years of schooling and European Universities. Proceedings of the 5th International Conference on Higher Education Advances (HEAd'19), Valencia (Spain), 1207-1216. https://doi.org/10.4995/HEAd19.2019.9494

Krull, G., & Duart, J. (2018). Meeting the Needs of Digital Learners: Learner Support Patterns and Strategies. In 13th International Conference on e-Learning (ICEL 2018), At Cape Town, South Africa.

European Commission (2010). EUROPE 2020 A strategy for smart, sustainable and inclusive growth, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52010DC2020&from=en

Hajian, S., Bonchi, F., & Castillo, C. (2016). Algorithmic Bias: From Discrimination Discovery to Fairness-aware Data Mining. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). Association for Computing Machinery, New York, NY, USA, 2125–2126. https://doi.org/10.1145/2939672.2945386

Marchisio, M., Rabellino, S., Roman, F., Sacchet, M., & Salusso, D. (2019). Boosting up Data Collection and Analysis to Learning Analytics in Open Online Contexts: an Assessment Methodology. Journal of E-Learning and Knowledge Society, 15(3), 49-59. https://doi.org/10.20368/1971-8829/1135048

Marquez-Vera, C., Morales, C. R., & Soto, S. V. (2013). Predicting School Failure and Dropout by Using Data Mining Techniques, in IEEE Revista Iberoamericana de Tecnologias del Aprendizaje, vol. 8, no. 1, pp. 7-14, Feb. 2013, https://doi.org/10.1109/RITA.2013.2244695

OECD/European Union (2019), Supporting Entrepreneurship and Innovation in Higher Education in Italy, OECD Skills Studies, OECD Publishing, Paris, https://doi.org/10.1787/43e88f48-en

Ochoa, X., & Merceron, A. (2018). Quantitative and Qualitative Analysis of the Learning Analytics and Knowledge Conference 2018. Journal of Learning Analytics, 5(3), 154–166. https://doi.org/10.18608/jla.2018.53.10

Riazy, S., Simbeck, K., & Schreck, V. (2020). Fairness in Learning Analytics: Student At-risk Prediction in Virtual Learning Environments. In Proceedings of the 12th International Conference on Computer Supported Education - Volume 1: CSEDU, ISBN 978-989-758-417-6, pages 15-25. https://doi.org/10.5220/0009324100150025

Scheffel, M., Drachsler, H., Stoyanov S., Specht, M.(2014). Quality Indicators for Learning Analytics. Educational Technology & Society, 17(4), 117–132.