Experimental validation of a massive educational service in a blended learning environment

Elena Baralis, Luca Cagliero, Laura Farinetti, Marco Mezzalama, Enrico Venuto
Politecnico di Torino
Torino, Italy
{elena.baralis, luca.cagliero, laura.farinetti, marco.mezzalama, enrico.venuto}@polito.it

***Abstract*—New information and communication technologies offer today many opportunities to improve the quality of educational services in universities and in particular they allow to design and implement innovative learning models. This paper describes and validates our university blended learning model, and specifically the massive educational video service that we offer to our students since 2010. In these years, we have gathered a huge amount of detailed data about the students' access to the service, and the paper describes a number of analyses that we carried out with these data. The common goal was to find out experimentally whether the main objectives of the educational video service we had in our mind when we designed it, namely appreciation, effectiveness and flexibility, were reflected by the users' behavior. We analyzed how many students used the service, for how many courses, and how many videos they accessed within a course (appreciation of the service). We analyzed the correlation between the use of the service and the performance of the students in terms of successful examination rate and average mark (effectiveness of the service). Finally, by using data mining techniques we profiled users according to their behavior while accessing the educational video service. We found out six different patterns that reflect different uses of the services matching different learning goals (flexibility of the service). The results of these analyses show the quality of the proposed blended learning model and the coherency of its implementation with respect to the design goals.**

I. Introduction

After many years of experience in distance education, in 2009 our university moved to a blended learning model [1] [2], where we introduced distance education elements to support the traditional presence model.

We intended to design a methodology where distance education elements could help and support different typologies of students, ranging from people that usually attend lectures in presence, to people that can participate only occasionally in live university activities: we have, in fact, a significant percentage of student workers. Besides the usual support made available by the teachers through the university educational portal (slides, solved exercises, exam texts and so on), we wanted to provide a most systemic intervention in this direction, trying to reach the best compromise between the educational effectiveness and the feasibility in terms of processes and costs.

The result was the decision to video-record in the classroom a significant number of courses (numbers are below in this section), and to make them available to students for video streaming or download from any kind of electronic device through the university educational portal. This solution has a number of advantages:

- it is familiar to students, which are used to the classroom context;
- it maintains a strong link between the presence and the distance activities, allowing their synchronization (all videos are available a few hours after recording thanks to a lean production process), so that a student that for any reason could not attend a live lecture can recover before the following one;
- it is flexible, by adapting to several level of "independence" from the live context (ranging from a complete synchronization to a complete self-adaptation of timings) so as to cover the needs of different typologies of students;
- it adapts very well to the new trend of users' preferences in terms of information access: videos are accessible via any kind of electronic device (computers, laptops, smartphones); a recent survey proposed to our students (about 6,000 responses were collected) demonstrated that smartphones are one of the favorite devices for accessing university services, and video is likely the most suitable and educationally effective content for smartphones' users;
- it is more cost-effective than other video recording solutions such as the TED model [3] (short talks in form of educational "pills" where the focus is on the quality of communication and on the incisiveness of the talk, and that therefore require significant investments in the production process).

We also noticed that this service improves the quality of participation in the live classrooms, because most "passive users" (students that in general participate to lectures with a passive attitude, never asking questions or making interventions, and rarely working on the proposed exercises) after a while preferred to follow the course remotely.

Our video courses are MOCs (Massive Online Courses) and intentionally not MOOCs (Massive Open Online Courses) [4]: in general, we provide access to our students only, because our model is not suitable for a larger audience. In fact, videos are the live recordings of a teacher in the classroom with minor post-processing, and therefore they are intrinsically tied to our educational context: the usability and the effectiveness for users outside this context is questionable.

However, we also experimented for several years a MOOC model for two courses, the ones we considered the most interesting for a larger audience: computer science and chemistry. In the case of computer science, for example, the thirty-nine videos got about 50,000 accesses every year by people outside our university community, and we consider this a good success.

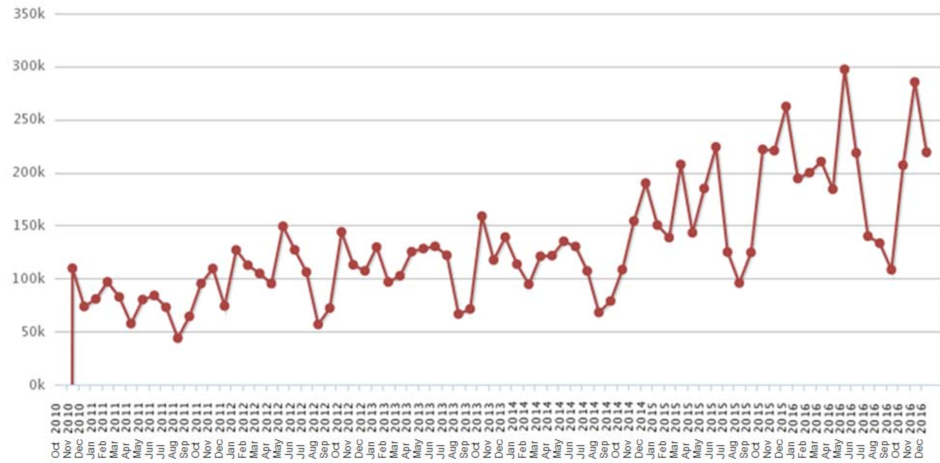

Fig. 1. Monthly accesses to the educational video service

The following sections will give some more details about the video educational service: Section A gives some concrete figures about the service, and Section B describes the characteristics of the educational video production and delivery processes.

A. *The numbers*

Since 2010 we have video-recorded in the classroom all the courses of the first year of the B.S. in Engineering (the first year is common to all B.S. engineering curricula), all the courses of the B.S. curricula in Computer Engineering, Electronic Engineering and Mechanical Engineering, and all the courses of the M.S. curriculum in Computer Engineering. Every year over sixty courses are fully live recorded, for about 3,000 videos available for streaming or download to the 15,000 students involved in these curricula, which represent more than 30% of the total number of students in our university. This generates about 1,200,000 video streaming/downloads per year.

Figure 1 shows the monthly accesses to the educational videos from the launch of the service (at the beginning of the 2010-2011 academic year) to the end of 2016. The graph reflects the cycles of activities within the academic years (the access peaks correspond to exam sessions) but it also positively shows a constantly growing trend, both within a single academic year and across academic years.

These videos, accessible through the university educational portal together with lots of other content, are a massive effort to support students in their learning process. The appreciation for this effort is tangible: at present, the number of logins per month exceeds 1,000,000 and the system provides access to about half a million of educational documents; in total, the number of downloads per year is over 10,000,000.

Besides, last six years showed an increasing diffusion of mobile devices, with a consequent higher and higher demand for mobile users' services. We followed this trend, by optimizing the educational services and content usability for any kind of device, from smartphones to powerful workstations, to offer a real multi-channel environment for education, which includes web applications as well as dedicated mobile Apps.

B. *Production and delivery*

The recording and the delivery of videos exploits a semi-automatic process, and the encoding and distribution environments use Open Source platforms and solutions.

Face-to-face teaching is video-recorded in the classroom, with a fixed camera operated by a technician (who typically is a part-time student with specific training) and a fixed background. The technician is in charge of the initial setup and of the supervision of the recording process (starts, endings, pauses); moreover, he or she tags the video with the appropriate content topic.

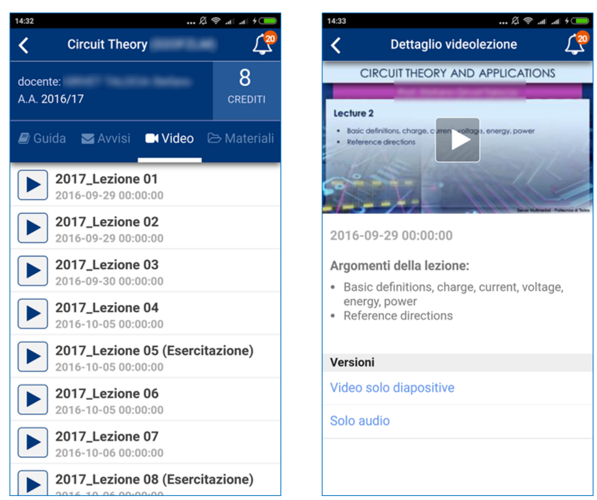

Fig. 2. Two screenshots of the educational video service on a mobile device: list of course videos (on the left) and video interface (on the right)

Classrooms are equipped with audio system, video-projector, computer, network connection, pen tablet (a touch-screen monitor that also acts as virtual blackboard) and codec. The adopted software environment for the teacher station is Open-Sankoré. Besides the video of the teacher, the multimedia flows coming from the tablet or other connected devices (as needed) are captured.

Lectures are processed as videoconferences and recorded on an IP VCR (IP-based videoconference recording, playback and streaming system). As soon as the lecture ends, the data flow transfer starts automatically: both the teacher's audio/video and the multimedia flow from the pen tablet are sent to a storage network. When the transfer is complete, the automatic editing phase starts, which inserts the two flows inside a template that contains the headings, a large frame for the slides and a smaller frame for the teacher. The teacher's face is included because it has a positive effect on students' response, mainly under the affective point of view [5].

The process results are a video file, a cover image with the list of lecture topics and an XML file that contains metadata such as the course name, the teacher, the academic year, the title of the lectures and so on. These three preliminary output files are automatically checked for consistency, and then the actual encoding phase starts, which generates several files in different video formats, optimized for a wide range of output devices, from powerful workstations to smartphones.

After encoding, videos are automatically published on the e-learning Open Source Chamilo LMS. Figure 2 shows two screenshots of the service on a mobile device.

II. RATIONALE AND CONTEXT OF THE RESEARCH

At the general level, the success of the designed blended learning model is demonstrated by the positive trend of the number of accesses and by the appreciation of the users collected via a number of questionnaires and interviews.

The designed model, however, has flexibility as one of its focuses, to cover the needs of as many different typologies of users as possible. We were interested therefore in understanding for what purposes students use the educational videos, in discovering different user profiles that represent significant categories of students, and in analyzing the actual educational effectiveness of the proposed service for the different profiles.

The analyses described in the following sections have the common goal to find out experimentally whether the main objectives of the educational video service we had in our mind when we designed it, namely appreciation, effectiveness and flexibility, were reflected by the users' behavior.

Learning analytics [6] are used to collect and measure data about learners and their context, to understand and optimize learning and the environment in which it occurs. The importance of applying learning analytics in video-based learning is well acknowledged [7] [8]. Many authors have worked in the direction of extending existing technological architectures with modules to support learning analytics, e.g. [9] and [10], or providing visual interfaces for visualizing learning analytics, such as [11] and [12].

However, the actual application of analytics is generally based on the measure of the learners' interaction through tools that complement the video-lessons and not on the videos; examples are performance in interactive quizzes [13], or participation to forums [14] or other social tools [15]. The reason is that most of video-based learning happens in a completely remote educational context, where providing students with effective ways for synchronous and asynchronous distance interaction is of fundamental importance. Our model is different, because we implement a blended learning environment, where interaction mostly happens in presence, and it is therefore difficult to track. Then, we need to apply our analyses on the data we systematically collect, i.e. video accesses.

Besides, the most popular video-based learning model today is based on MOOCs, and most of the studies on learning analytics are relative to this context, for example [16] and [17]. Our educational context, however, is very different from MOOCs: we have a more controlled environment, with homogeneous users in terms of learning pre-requisites, and very detailed collected data. Besides, the most important goal for collecting and analyzing data is very different too: in the case of MOOCs, the most important concern is to understand the reasons for drop-offs for limiting their occurrence (see [18] and [19] as examples). Our main concern, on the contrary, is to evaluate learning effectiveness and flexibility of use. Specifically, we would like to find answers to a number of questions:

1. How many students use the educational video service? For how many courses? In a specific course, are there videos that have a higher number of accesses than others, and why?
2. Is there a positive correlation between the use of the educational video service and the students' performance, in terms of exam success rate and average mark?
3. Is it possible to extract significant patterns of students' behavior when accessing the videos? Do they reflect specific learning goals?
4. Do students develop and apply coherent learning strategies for different courses, about the use of the educational video service?

Since the launch of the educational video service in 2010 we have collected a huge amount of very detailed data about students' accesses. We can extract information about when a student access a specific video-lecture, via what kind of device, how many times the same students accesses the same video-lecture and so on. Data are relative to several academic years and a large number of courses (more than one hundred) regarding different branches of Engineering (both at the B.S. and at the M.S. levels), for which the number of involved students varies from a few dozen up to several thousands.

To work with suitable and coherent data, we decided to concentrate the analyses reported in the paper to the specific case of the compulsory courses of the first year of all the B.S. in Engineering. These courses are only five: Computer Science, Calculus and Chemistry in the fall semester, Physics and Geometry in the spring semester, but the number of students that have these courses in their curricula exceeds 9,000 every year. Besides, the analyses will consider only the specific academic year 2015-2016 (from October 2015 to September 2016). In this way, we defined a clear context with significant statistical data.

Each of the following sections focuses on one of these questions, by mining the collected data in search for suitable answers.

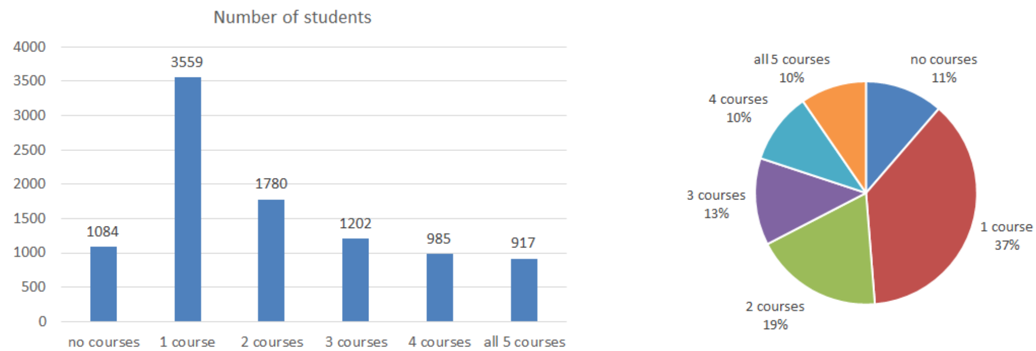

Fig. 3. Number of courses accessed by the students – absolute and relative values

III. Question 1: Access to Videos

For all the 9,527 students that have the courses of the first year of the B.S. in Engineering in their curriculum (the first year is common to all branches of Engineering) we analyzed the number of (compulsory) courses for which, during the academic year 2015-2016, they used the educational video service. The total number of these courses is five, and Figure 3 shows the results of this analysis in absolute value and in percentage. Figure 3 reports, for example, that 3,559 students out of 9,527 used the service only for one course, and ignored it for the other four. For "using the service" here we consider if the student accessed at least one of the videos of the course. The data of the students were anonymized (we used the MD5 value of the student's identification number piped with a secret passphrase), maintaining however the link between their access records and their performance records, to be able to explore also the correlation between these two aspects.

The graph shows that only a small percentage of students (11%) ignored the service, and did not access any of the recorded courses. Most of the students (37%) used the service only for one specific course, and more than half of the students (56%) used it for a small number of courses, one or two.

The interpretation of these data, supported also later by the analysis in Section V, is that videos demonstrate to be mainly a tool for supporting students that look for an extra help when necessary. This happens in the case of specific courses where they experience more difficulties in understanding concepts or in applying theory to practice, or (more prosaically) when they fail to pass the exam.

Then, we analyzed the access to each of the videos that make up a course, for the five compulsory courses of the first year of all the B.S. in Engineering: Computer Science, Calculus, Chemistry (fall semester), Physics and Geometry (spring semester) in a period corresponding to the academic year 2015-2016 (from October 2015 to September 2016).

Figure 4 shows the result of this analysis for two of these courses, Computer Science and Physics. The graphs report the average access rate for each single video (39 in the case of Computer Science and 51 in the case of Physics) of the courses. The number of videos per course varies: the lectures are video-recorded in real time and consequently their lengths are different, depending on the covered topics and the possible pauses made by the teacher. The average access rate is the total number of accesses made by the students to the videos of the course, divided by the total number of students that have the course in their curriculum (9,527 students). In the graphs, the dark bars identify the videos tagged as "theory" by the teachers, and the light ones the videos tagged as "practice".

Figure 4 shows that the number of accesses is very high in general: on average students accessed almost every video at least twice. In reality, since not every student is an "active" user of the educational video service (this aspect will be analyzed specifically in Section V, but we can anticipate that for example in the case of Computer Science 59% of the students did not access a single video), the actual access rate for the videos is more than double.

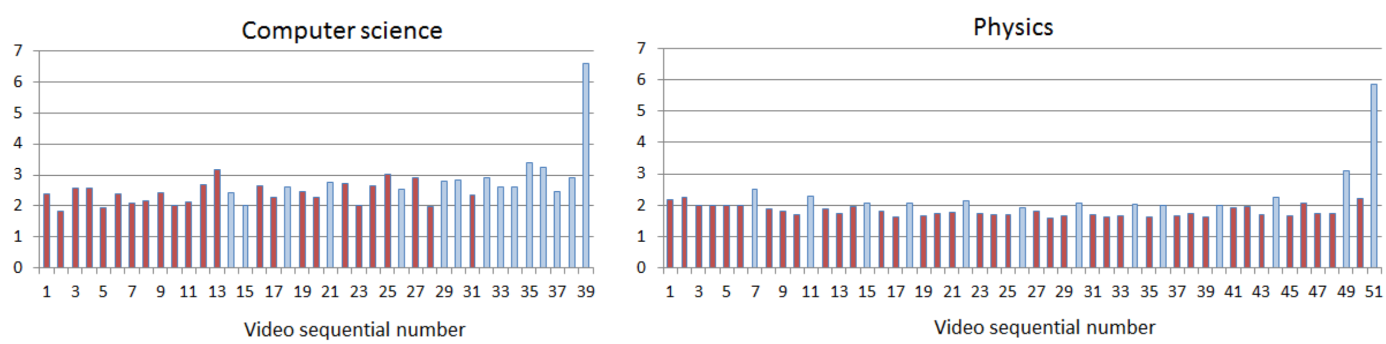

Fig. 4. Average number of accesses per student to each video of the computer science course (on the left) and of the physics course (on the right)

From these graphs, we can understand that our users acknowledge the usefulness of the educational videos, especially the ones with a "practice" content, and that positively the number of accesses they make does not depend on the progressive number of the video.

IV. QUESTION 2: CORRELATION BETWEEN ACCESS AND PERFORMANCE

In this analysis, our goal was to find the correlation between the use of the educational video and the students' performance. We considered each of the five compulsory courses of the first year of the B.S. in Engineering separately, and we report in the paper the data about the course of Computer Science, selected as the most representative since it is the only one that has been recorded in the reference academic year (2015-2016).

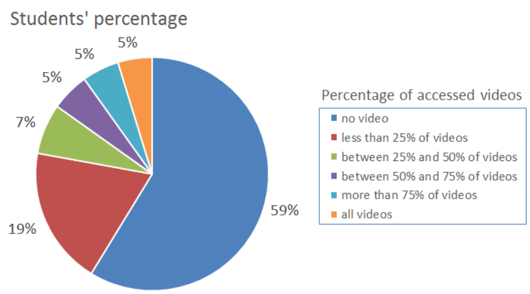

Fig. 5. Percentage of computer science videos accessed by the students

We divided the students into six categories, depending on the number of videos they accessed with respect to the total number of the videos of the course, which is 39 in the case of Computer Science. The six categories are: no access to any video, access to less than 25% of the videos, between 25% and 50%, between 50% and 75%, more than 75% and access to all videos. The total number of students is 9,527.

Figure 5 shows the percentage of students in each category. More than half of the students (59%) did not use the videos at all for the Computer Science course. Another 19% accessed only a small number of videos (less than 25%), and only 15% of the students accessed more than half of the videos.

Figure 6 and Figure 7 analyze two different aspects of students' performance: the exam success percentage and the average mark. The success percentage is the number of students in each category that succeeded to pass the exam in one of the four sessions that took place in the 2015-2016 academic year, divided by the total number of students that belong to the same category. The bars in Figure 6 show the number of students in each category, and the line their success percentage. The categories with the best performance under this criterion are the ones that accessed a small number of videos (especially the category less than 25%). The line shows that in general the performance of the students that used the videos is higher than the performance of the students that did not, with the exception of the last category (students that accessed all the videos). This last outcome is justified by the fact that we can monitor the access to the videos but not their actual "use" by the students. Since accesses include video streaming and downloads, it is very likely (and supported by the case studies in Section VI) that a large number of students in the last category simply downloaded all the videos to keep them for future use.

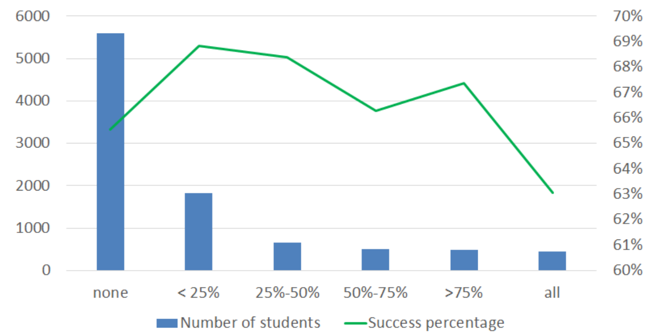

Fig. 6. Correlation with exam success rate

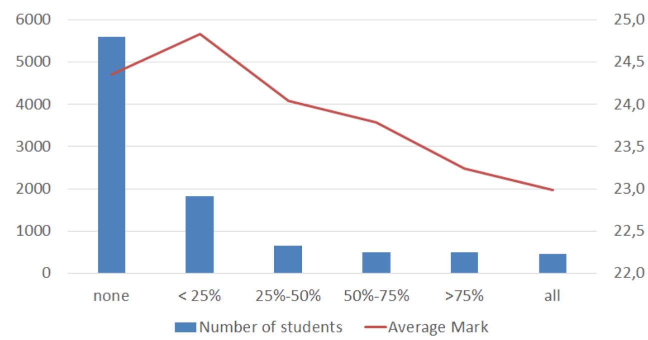

Fig. 7. Correlation with average mark

The line in Figure 7 shows the average examination mark for each of the categories, which in our university system is an integer number between 18 and 30. The students' performance under this criterion has a similar pattern: again, the best category is the one of the students that accessed a small number of videos, and the worst is the one of the students that accessed all of them. This graph clearly shows than a large number of students are confident on what they learned in the classroom and do not consider the use of video-lectures necessary, and that their performance is quite good. However, the best students are the ones that use videos when necessary, for example to review specific topics that were not clear enough, or to practice again specific exercises, or in case they were not able to attend a specific lecture.

Considering the two performance indicators together, we have a positive correlation between the use of the videos and the chance of passing the exam, suggesting that their main role is not to substitute live lectures but to complement them, supporting students when they need to fill a gap. Positively, this role is very coherent with the rationale for which the service has been introduced in our university: to provide extra support for students in a blended learning environment.

V. QUESTION 3: STUDENTS' PROFILING

To profile the usage of the educational video service we analyzed the accesses of the students to the videos separately for each course. Specifically, for each student we considered the number of accessed videos, the dates of the first and the last access to any video of the course, and the corresponding temporal gap (i.e., the difference between the last and the first access) expressed in days. To cluster students according to their service usage pattern, we graphically analyzed the pairwise distances between the corresponding records and we identified six groups of highly similar records. Specifically, we measured the pairwise distance between records by using the mixed Euclidean distance available in the machine learning and data mining Rapid Miner tool (https://rapidminer.com).

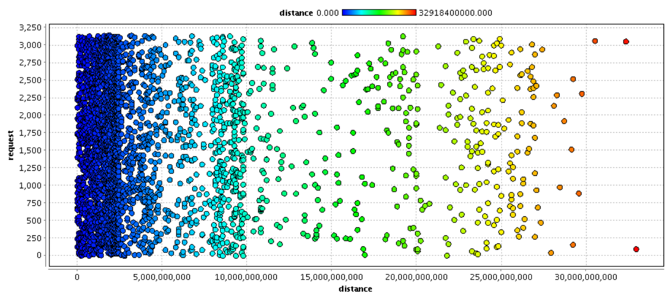

Fig. 8. Scatter plot used to identify clusters of students

To identify the borders of the clusters, we first selected the centroid of a dense regions of records and then we used the scatter plot to set the maximal intra-cluster distance between records. Figure 8 shows an example of scatter plot, where the distance from the considered centroid is on the x-axis, while the identifiers of the candidate neighbors (i.e., the student id) are on the y-axis. Students with a pairwise distance lower than a specified threshold (10^9 in our experiments) are included in the same cluster because they show similar behavior in terms of the educational video service usage. In our analysis, we disregarded the subset of students who never accessed any video (and consequently the total number of considered students is 3,143).

We used the six clusters of students for user profiling. In the next subsections, we discuss the six clusters; each of them is relative to one of the user profiles and it is exemplified by the graph of a specific representative student. Each of the user profiles has a name, which outlines its peculiar characteristics (e.g. "synchronous user" or "exam-driven user"). Table I summarizes the six clusters, highlighting the usage type of the educational video service and the main attitude of the student. The table also reports the coverage of the cluster in our data set, i.e. the percentage of students that belong to the cluster, and whether the usage of the educational video service was effective, i.e. if the students succeeded in passing the exam. A small percentage of students (about 3%) does not belong to any of the selected clusters since they show an anomalous behavior.

To make the exemplificative graphs easily comparable, the following subsections consider the computer science course only. In the graphs, the x-axis represents the dates on which a video was accessed by the user (dates are relative to the whole 2015-2016 academic year); note that the computer science course is given in the fall semester, i.e. in the first semester of the academic year. The y-axis represents the video sequential number, which in the case of computer science is a number between 1 and 39. The blue dots, connected by the blue lines, represent the event of a specific video access.

A. *Cluster A: "synchronous user"*

The first graph, reported in Figure 9, shows a student that accessed all videos mainly in sequence, during the whole fall semester and practically in parallel with the live lectures. Sporadically he or she came back to review a previous video, and when the exam was close the student increased the activity in particular accessing several times the last videos, that contain many exercises in preparation for the exam. This student passed the exam in February 2016, during the first exam session.

This graph represents a student profile with a systematic attitude that effectively use videos as the main study tool.

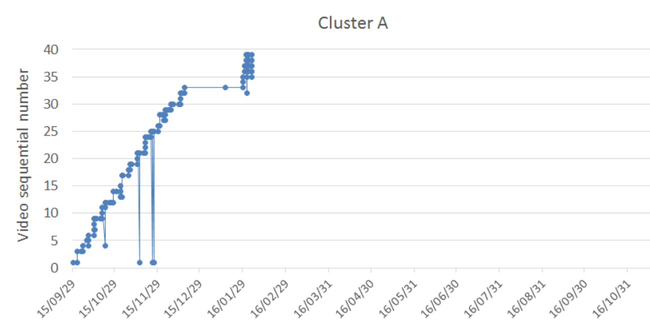

Fig. 9. Cluster A: video accesses by a representative student

TABLE I. SUMMARY OF CLUSTER CHARACTERISTICS

Cluster	Name of profile	Description	Video usage	Attitude	Success	Coverage
A	Synchronous user	Students that access all videos, in parallel with live lectures	Synchronous study	Systematic	Y	39%
B	Just-enough user	Students that look for specific information when necessary; typically, they do not access all videos	Review	Selective	Y	21%
C	Exam-driven user	Students that concentrate video accesses in very narrow periods, typically close to the exam sessions	Squeezed study	Superficial	N	15%
D	Asynchronous user	Students that access all videos, with different timing w.r.t. live lectures	Asynchronous study	Independent	Y	4%
E	Focused user	Students that access all videos within a short period, typically after an exam failure	Failure recovery	Motivated	Y	6%
F	Drop-out user	Students that access only the first videos and then quit	Incomplete study	Unmotivated	N	12%

B. *Cluster B: "just-enough user"*

The second graph, reported in Figure 10, shows a student that also accessed all videos, but that concentrated his or her accesses in three different periods. The graph shows that he or she did not reach a complete preparation for the first exam session in February (he or she did not access all the videos), and in fact he or she did not participate to it. The student then considered starting again the study of computer science during the spring semester, but he or she gave up quickly, and again he or she did not participate to the second exam session, in July. Finally, the students decided to study the course seriously during the summer, and in fact, he or she passed the exam in the third session, in September.

This graph represents a student profile that effectively uses videos as tool for reviewing concepts and practicing in preparation for the exam sessions, with a selective attitude, i.e. using the videos that he or she considers the most suitable for his or her needs.

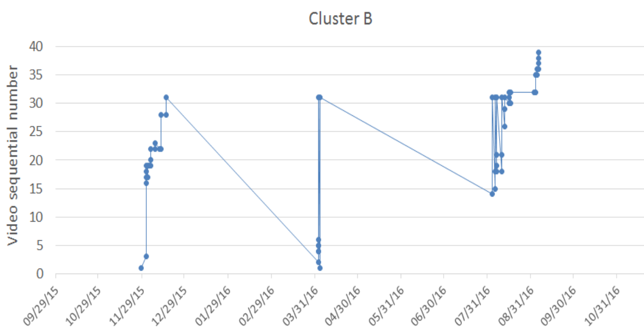

Fig. 10. Cluster B: video accesses by a representative student

C. *Cluster C: "exam-driven user"*

The third graph, reported in Figure 11, shows a student that apparently has a behavior similar to cluster B: he or she accessed all videos, and concentrated his or her accesses in three different periods. In this case, however, the periods are very close to the three exam sessions (February, July and September), and moreover the accesses happened in very few days; this situation very likely means that they represent video downloads and not streaming, and therefore it is possible that the videos were not actually used. The student did not participate in the February exam session, failed the July one and withdrew in the September one.

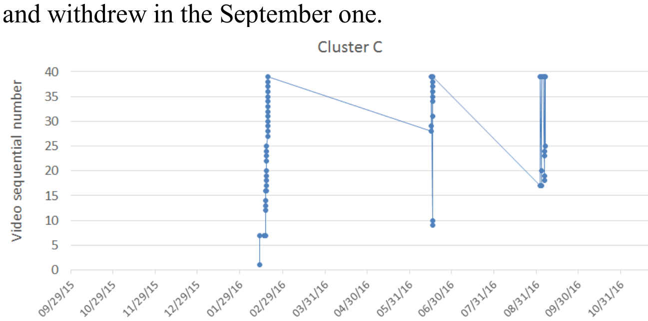

Fig. 11. Cluster C: video accesses by a representative student

This graph represents a student profile that acknowledges the importance of the videos as a study and review tool, but has a superficial attitude. This pattern represents a non-effective use of the educational video service.

D. *Cluster D: "asynchronous user"*

The fourth graph, reported in Figure 12, shows a student that accessed all videos mainly in sequence, but during the spring semester, i.e. not in parallel with the live lectures (that are in the fall semester). This student passed the exam in July 2016, as soon as he completed the video course.

This graph represents a student pattern with an independent attitude that use videos as the main study tool, but in a different way with respect to cluster A: non-simultaneously with the live lectures. For this pattern, videos are an essential tool to catch up with the exams that were left behind.

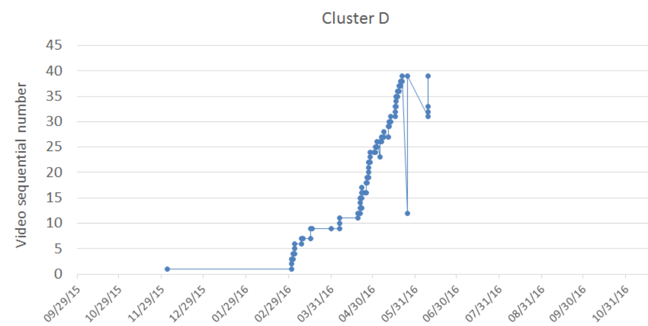

Fig. 12. Cluster D: video accesses by a representative student

E. *Cluster E: "focused user"*

The fifth graph, reported in Figure 13, shows a student that accessed all videos in about a month period, during the February exam session. This exam session has two different exam dates, one at the beginning and the other at the end of the month; students have the possibility to participate in both of them, if necessary. This student participated in the first exam test without accessing any video, and failed it. He or she then used the educational video service and passed the exam in the second exam date of February.

This graph represents a student profile that effectively uses videos as a failure recovery tool, as soon as he or she understands (by failing an exam) that the preparation is insufficient and that the he or she needs extra support. These users are generally very motivated.

Fig. 13. Cluster E: video accesses by a representative student

F. *Cluster F: "drop-out user"*

The sixth graph, reported in Figure 14, shows a student that accessed less than half of the videos: he or she never went beyond lecture 16 (out of 39). At the beginning of the fall semester, this student accessed the videos mainly in sequence, practically in parallel with the live lectures; however, approximately at the middle of the course he or she gave up. The student made very little effort to recover for the February and July exam sessions (he or she failed the February exam session), but he or she started again with the exam preparation during the summer break. Unfortunately, the student gave up again, and he or she did not pass the exam in the 2015-2016 academic year.

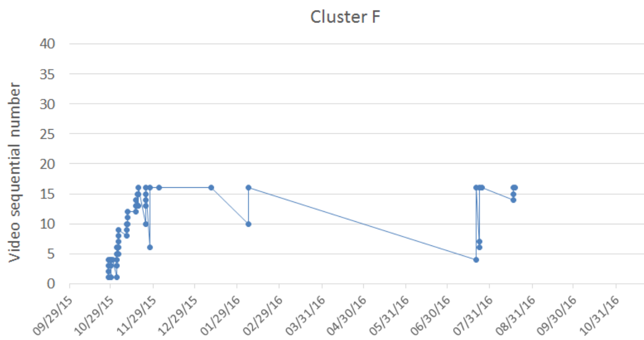

Fig. 14. Cluster F: video accesses by a representative student

This graph represents a student profile that acknowledges the importance of the videos as a study tool, and tries to use them seriously for a systematic preparation, but that for some reason is not able to follow this strategy to the end. Possible factors that prevented success are lack of motivation, or difficulty in organizing and managing the learning process when the synchronous link with the live lectures is broken. This profile represents a non-effective use of the video service.

G. Outlier user

For the sake of completeness, we also report in Figure 15 the graph of a student that does not belong to any of the identified clusters. It shows a student that accessed all videos, but concentrating all the accesses in just three days during the summer: except possibly the first one, the other accesses are obviously downloads.

This graph represents a student that uses videos during the summer university break to catch up with the exams that he or she left behind. The fact that he or she accessed the videos exactly in mid-August make his or her behavior quite peculiar, and this is confirmed by a very small number of students (less than 1%) that share the same profile.

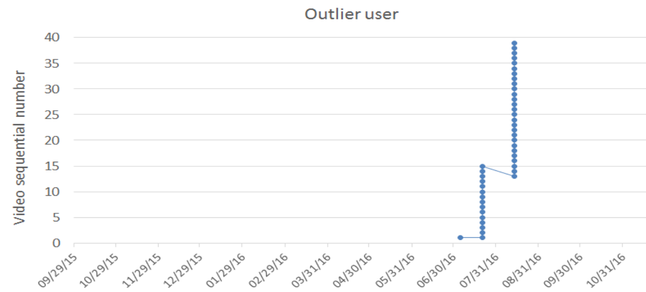

Fig. 15. Outlier user: video accesses

H. Analysis of the six clusters

The six clusters outline six different profiles of video users, some of which are successful and some not. To summarize, the main "educational" roles of videos, according to the students' behavior, are:

- Study a course during the whole semester: cluster A (successful) and cluster F (unsuccessful).
- Review concepts and practice in preparation to an exam session: cluster B (successful) and cluster C (unsuccessful).
- Catch up with the exams left behind: cluster D (successful) and outlier user (unsuccessful).
- Recover when necessary, for example after failing an exam: cluster E (successful).

The educational video service, then, demonstrates to be flexible enough to accommodate different users' needs, and distinct learning models, where distance education merges with presence education in different level of balance. The actual success of the students, of course, depends on the constant and coherent use of the educational tools.

The coverage numbers in Table I show that the most represented cluster is the one of the synchronous users, and that the cluster of asynchronous users has a much lower percentage of students. This suggests that the link between live lectures and videos is very strong, and most of the students prefer a blended learning model.

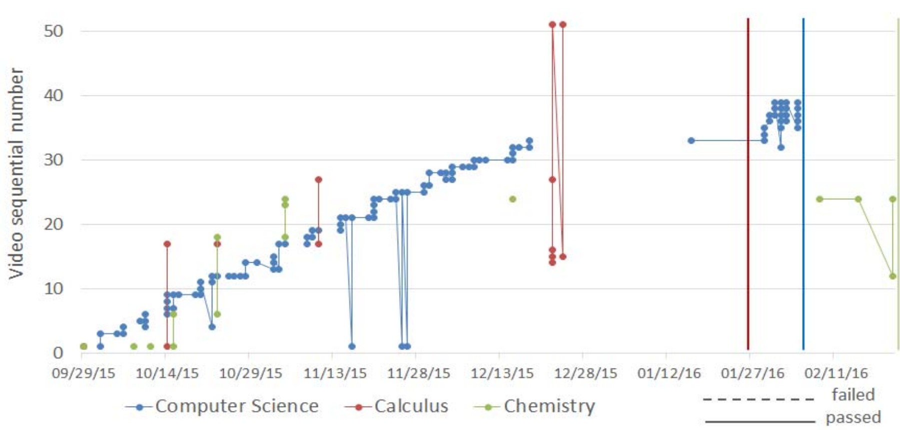

Fig. 16. Student representing cluster A: access to all videos and performance during the fall semester

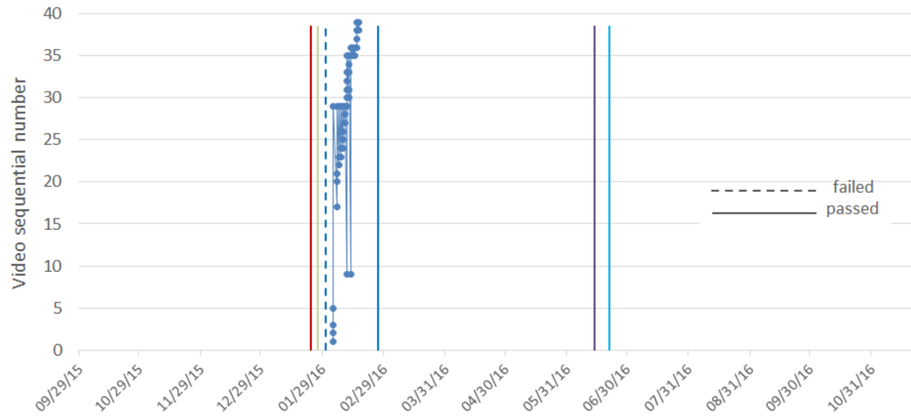

Fig. 17. Student representing cluster E: access to all videos and performance during the entire academic year

Besides, the "just-enough" user profile is also very well represented: this category of students (as analyzed in Section IV) is the one with the best performance in terms of success percentage and average mark.

Finally, we can positively outline that the "successful" clusters (A, B, D and E), that represent an effective use of the educational video service, cover at least 70% of the users.

VI. QUESTION 4: STUDENTS' LEARNING STRATEGIES

Finally, we analyzed more in depth the behavior of two of the students considered in Section V, and specifically the student that represents cluster A and the student that represents cluster E because they are successful patterns with different learning objectives. The objective of this analysis was to understand whether the behavior of the students was constant throughout the different courses, or the learning goals were different in different courses.

A. Case study 1: student representing cluster A

The graph reported in Figure 16 shows the behavior of the student during the fall semester, where he or she had to follow three compulsory courses: Computer Science, Calculus and Chemistry. As in the previous graphs, the dots represent the event of a specific video access; the number of videos is not constant for all the courses: Computer Science has 39 lectures, Calculus 54 and Chemistry 51. The vertical lines represent the event related to exams: the color refers to the course, and the continuous line shows a passed exam while a dashed line a failed one (in this specific graph there are no exam failures). We already discussed the behavior of the student in the computer science course (systematic study and success at the first exam session, see section V.A). For the other two courses, the student used a different strategy: he or she accessed only a few videos when needed, probably to review specific concepts that were unclear or to practice specific exercises.

For each of the courses, the student demonstrated a coherent attitude in working throughout the whole semester; this strategy obviously pays because he or she passed all the exams in the first available session. The different behavior is probably due to a different self-confidence level about the course topics: in case of computer science, he or she clearly felt the need for more support.

B. Case study 2: student representing cluster E

The graph reported in Figure 17 shows the behavior of the student during the entire academic year, where he or she had to follow all the five compulsory courses: Computer Science, Calculus, Chemistry, Physics and Geometry. From the graph, we can see that the student used the educational video service only for Computer Science, after that he or she failed the first exam attempt (we already discussed this in section V.E). For all the other courses, the students had no problem in passing the exam at the first attempt, and therefore he or she has never used the corresponding video courses.

This case shows a coherent behavior, and the role of the educational video service, for this student, is to help to recover just in case of demonstrated failure.

VII. CONCLUSIONS AND FUTURE WORK

Through the analysis of the available data about students' accesses to the educational video service, we gave an answer to the proposed questions, and specifically:

1. *How many students use the educational video service? For how many courses? In a specific course, are there videos that have a higher number of accesses than others, and why?*
   The large majority of students uses the educational video service at least for one course; most of the students use it for a small number of courses (one or two). This implies on the one hand that students recognize the value of the service, and on the other hand that they use videos only when necessary: the service then complies with the "just in case" learning paradigm.
   Besides, the number of accesses does not depend on the progressive number of the video, but rather on its content: the drop-off rate is very low, and students tends to access more frequently the videos that contain practice rather than theory. This implies on the one hand that students acknowledge the quality of the service by using it from the start to the end, and on the other end that the videos are able to satisfy the specific need of applying theory into practice.
2. *Is there a positive correlation between the use of the educational video service and the students'*

performance, in terms of exam success rate and average mark?

Yes, there is a positive correlation: the probability of passing the exam is much higher for the students that use the educational video service; besides, the category of students that has the major benefit in term of success rate and average mark is the one that uses a small percentage of videos. This implies on the one hand that the educational video service is a very helpful tool, especially to reach an adequate level of competence (sufficient to pass the exam). On the other hand, this implies that the most effective scenario is again the "just in case" one: students with the best performance are the ones that access specific videos when they need them.

3. *Is it possible to extract significant patterns of students' behavior when accessing the videos? Do they reflect specific learning goals?*

   Yes, we could extract from data six different students' patterns, which use the educational videos for different purposes: study in a blended learning situation, study in a completely remote learning situation, review in preparation for the exam, and recover from fault when necessary. This implies that the service has a very good level of flexibility, by adapting to different students' learning goals and to different typology of users, ranging from full-time students to student workers.

4. *Do students develop and apply coherent learning strategies for different courses, about the use of the educational video service?*

   The preliminary analysis we made, with a small number of "performant" students, showed that students tend to develop a personal strategy for the effective use of the service, which remains coherent throughout all the courses. This implies that students are able to adapt a service designed to be flexible to their own goals and needs.

These answers represent an experimental proof of the quality of the educational video service under different points of view, namely acceptance, effectiveness, and flexibility. In particular, the idea of profiling users according to their behavior in accessing educational videos demonstrated to be a way for validating "ex-post" our blended learning model especially in terms of actual flexibility.

Future work will focus on profiling students and courses by applying unsupervised data mining techniques, such as clustering and association rule mining. The goal is to capture the most interesting patterns in the analyzed data by exploiting automatic techniques, which are able to scale towards very large datasets.

REFERENCES

[1] J. Stein, and C. R. Graham, "Essentials for blended learning: a standards-based guide", New York, NY: Routledge, 2014.

[2] F. Alonso, D. Manrique, L.Martínez, and J. M. Viñes, "How blended learning reduces underachievement in higher education: an experience in teaching computer sciences", in IEEE Transactions on Education, vol. 54, no. 3, August 2011, pp. 471-478.

[3] TED ed lessons worth sharing, https://ed.ted.com/

[4] J. Kay, P. Reimann, E. Diebold, and B. Kummerfeld, "MOOCs: so many learners, so much potential ...", in IEEE Intelligent Systems, vol. 28, no. 3, May 2013, pp. 70-77.

[5] R. F. Kizilcec, K. Papadopoulos, and L. Sritanyaratana, "Showing face in video instruction: effects on information retention, visual attention, and affect", in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI 2014*, April 26–May 1, 2014, Toronto, Ontario, Canada.

[6] R. Ferguson, "Learning analytics: Drivers, developments and challenges," in International Journal of Technology Enhanced Learning, vol. 4, No. 5/6, January 2012, pp. 304–317.

[7] M. N. Giannakos, K. Chorianopoulos, M. Ronchetti, P. Szegedi, and S. D. Teasley, "Analytics on video-based learning", in *Proceedings of the Third International Conference on Learning Analytics and Knowledge LAK '13*, April 8-12, 2013, Leuven, Belgium.

[8] M. N. Giannakos, D. G. Sampson, and Łukasz Kidziński, "Introduction to smart learning analytics: foundations and developments in video-based learning", in Smart Learning Environments 2016, 3:12.

[9] S. Boyer, B. U. Gelman, B. Schreck, and K. Veeramachaneni, "Data science foundry for MOOCs", in *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, October 19-21, 2015, Paris, France.

[10] J. A. Ruipérez-Valiente, P. J. Muñoz-Merino, and C. Delgado Kloos, "An architecture for extending the learning analytics support in the Khan Academy framework", in *Proceedings of the First International Conference on Technological Ecosystem for Enhancing Multiculturality TEEM '13*, Salamanca, Spain, November 14-15, 2013, pp. 277-284.

[11] H. Qu, and Q. Chen, "Visual analytics for MOOC data", in IEEE Computer Graphics and Applications, Vol. 35, No. 6, November-December 2015, pp. 69-75.

[12] M. N. Giannakos, K. Chorianopoulos, and N. Chrisochoides, "Collecting and making sense of video learning analytics", in *Proceedings of the 2014 IEEE Frontiers in Education Conference (FIE)*, October 22-25, 2014, Madrid, Spain.

[13] C. G. Brinton, S. Buccapatnam, M. Chiang, and H. V. Poor, "Mining MOOC clickstreams: video-watching behavior vs. in-video quiz performance", in IEEE Transactions on Signal Processing, Vol. 64, No. 14, 2016, pp. 3677 – 3692.

[14] C. G. Brinton, M. Chiang, S. Jain, H. Lam, Z. Liu, and F. M. Fai Wong, "Learning about Social Learning in MOOCs: From Statistical Analysis to Generative Model", in IEEE Transactions on Learning Technologies, Vol. 7, No. 4, October-December 2014, pp. 346-359.

[15] C. Alario-Hoyos, M. Pérez-Sanagustín, C. Delgado-Kloos, H. A. Parada G., and M. Muñoz-Organero, "Delving into participants' profiles and use of social tools in MOOCs", in IEEE Transactions on Learning Technologies, Vol. 7, No. 3, July-September 2014, pp. 260-266.

[16] Q. Chen, Y. Chen, D. Liu, C. Shi, Y. Wu, and H. Qu, "PeakVizor: visual analytics of peaks in video clickstreams from massive open online courses", in IEEE Transactions on Visualization and Computer Graphics, Vol. 22, No. 10, 2016, pp. 2315-2330.

[17] M. A. Mercado-Varela, A. García-Holgado, F. J. García-Peñalvo, and M. S. Ramírez-Montoya, "Analyzing navigation logs in MOOC: a case study", in *Proceedings of the Fourth International Conference on Technological Ecosystem for Enhancing Multiculturality TEEM '16*, Salamanca, Spain, November 2-4, 2016.

[18] M. Fei, and D. Y. Yeung, "Temporal models for predicting student dropout in massive open online courses", in *Proceedings of the 2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, November 14-17, 2015, Atlantic City, NJ, USA, pp. 256-263.

[19] R. Ferguson, and D. Clow, "Examining engagement: analysing learner subpopulations in massive open online courses (MOOCs)", in *Proceedings of the 5th International Learning Analytics and Knowledge Conference (LAK15)*, March 16–20, 2015, Poughkeepsie, NY, USA.