PRECISE EFFECTIVENESS STRATEGY FOR ANALYZING THE EFFECTIVENESS OF STUDENTS WITH EDUCATIONAL RESOURCES AND ACTIVITIES IN MOOCS

Pedro J. Muñoz-Merino[a], José A. Ruipérez-Valiente[a,b], Carlos Alario-Hoyos[a], Mar Pérez-Sanagustín[a], Carlos Delgado Kloos[a]

[a]Universidad Carlos III de Madrid, Av. Universidad 30, 28911 Leganés (Madrid), Spain

[b]IMDEA Networks Institute, Av. del Mar Mediterráneo 22, 28918 Leganés (Madrid), Spain

* Corresponding author info: Pedro J. Muñoz-Merino (phone: (+34) 916246233, email: pedmume@it.uc3m.es)

ABSTRACT

Present MOOC and SPOC platforms do not provide teachers with precise metrics that represent the effectiveness of students with educational resources and activities. This work proposes the Precise Effectiveness Strategy (PES), a generic methodology for defining precise metrics that enable the calculation of the effectiveness of students when interacting with educational resources and activities in MOOCs and SPOCs, taking into account the particular aspects of the learning context. PES has been applied in a case study, calculating the effectiveness of students when watching video lectures and solving parametric exercises in four SPOCs deployed in the Khan Academy platform. Different within courses and between courses visualizations are presented combining the metrics defined following PES. We show how these visualizations can help teachers make quick and informed decisions in our case study, enabling the whole comparison of a large number of students at a glance, and a quick comparison of the four SPOCs divided by videos and exercises. Also, the metrics can help teachers know the relationship of effectiveness with different behavior patterns. For example, results from using PES in the case study revealed that the effectiveness metrics proposed were moderately negatively correlated with some behavior patterns like recommender listener or video avoider.

Keywords

Learning analytics, MOOCs, SPOCs, Precise Effective Strategy, metrics.

1. INTRODUCTION

Massive Open Online Courses (MOOCs) have brought about a revolution in education [1]. During the last year many teachers and higher education institutions joined the MOOC wave, launching courses in different areas. Platforms like Coursera, edX, MiríadaX, or FutureLearn are providing the support teachers need to deploy open courses that may scale up to thousands of participants.

Most of the MOOCs deployed in the aforementioned platforms follow a common structure in which educational resources are offered as short video lectures, accompanied by activities (e.g., automatic correction exercises) that cover both formative and summative assessment activities [2]. These types of MOOCs, also known as xMOOCs, are the most widespread, and follow the so-called broadcast model [3].

The impact of MOOCs goes beyond providing free and open education to students worldwide, and is now leading to new blended learning scenarios at schools and Universities in which the main ideas of MOOCs are exploited to enhance teaching and learning in the form of successful flipped classrooms (i.e. students watch videos with the theoretical concepts from home and practice these concepts with automatic correction exercises, and later attend to the classroom to solve problems with teachers) [4]. Such use of the affordances that emerge from MOOCs to improve the quality of teaching and learning in traditional educational settings leads to what has been called SPOCs (Small Private Online Courses) in the media [5]. For example, Harvard University (through its edX brand HarvardX) has already taken a step forward, launching SPOCs for their Design school and Law school students [6].

In these educational contexts in which there can be a very large number of students, teachers need precise strategies to know what is going on with each individual student and with the whole class. Current Learning Analytics (LA) techniques [7] enable to collect huge amounts of low-level data regarding for instance students' interaction among themselves or with educational resources and activities, both in MOOCs and SPOCs [8]. From these low-level data it is possible to infer higher level behaviors and metrics that can be presented to the teacher as simple and understandable visualizations [9].

Assuming that the objective pursued by teachers in both MOOCs and SPOCs is that students complete all the proposed contents in a correct way (both educational resources such as video lectures, and activities such as automatic correction exercises), then it is necessary to propose metrics to determine how effective students were with respect to this objective, in other words, metrics to determine *the effectiveness of students with educational resources and activities*. These metrics can facilitate the classification of students according to their degree of effectiveness in a course, and the comparison of the overall effectiveness of the class in MOOCs and SPOCs using quantitative measures.

Nevertheless, as far as we know, the existing metrics that capture students' interactions with educational resources and activities in MOOCs and SPOCs are very rough (e.g., if a video was completed, number of exercises accessed and completed…). Further, these metrics do not take into account how educational resources and activities were structured (e.g., the most important parts in a video lecture), and how they relate to each other (e.g., the suggested order of completion). Visualizations to see at a glance the effectiveness of thousands of students are also missing in current platforms. For instance, the Khan Academy, which offers one of the most detailed learning analytics, only represents with colors four states of an activity (proficiency, struggling, started or not accessed), and does not provide graphical visualizations to represent how students' progress in a video.

In this paper, we present the Precise Effectiveness Strategy (PES), which is a generic methodology for supporting the calculation of the effectiveness of students when interacting with educational resources and activities in both MOOCs and SPOCs. PES guides the process for defining precise metrics to calculate this effectiveness, considering the main particularities of the learning context (e.g. the relationships among the different video resourceds). PES has been applied to a case study to calculate the effectiveness of students when interacting with video lectures and automatic correction exercises in four remedial SPOCs deployed in the Khan Academy platform. The metrics defined in the case study enabled: 1) to represent the effectiveness of students' interactions with videos and exercises in a simple way; 2) to measure the relationship of the effectiveness metrics with behavior patterns; and 3) to compare the overall effectiveness of students when interacting with educational resources and activities between the four SPOCs.

The remainder of this paper proceeds with a review of the literature regarding the definition of effectiveness in educational contexts in Section 2. Section 3 presents PES, detailing its four phases. Then, Section 4 describes the case study and the specific metrics proposed for it. Section 5 presents and discusses the results obtained from the case study. Finally, Section 6 draws the conclusions and summarizes the main lines of future work.

2. RELATED WORK

Much of the traditional educational literature addresses the concept of effectiveness from the perspective of learning (*students' learning effectiveness*): "how much did the students learned, how well did they master skills and how well can they apply knowledge" [10]. The concept of effectiveness applies to face-to-face, blended and online education, but becomes more important in the latter, where teachers cannot easily track learning gains [11, 12]. In order to measure learning effectiveness, most authors typically use (if possible) achievement tests or surveys for collecting student perceptions [13].

The study by Swan [11] goes further and proposes measuring students' learning effectiveness in terms of interactivity with peers (social presence), with instructors (teaching presence) and with contents (cognitive presence) [14]. Following this idea, it is possible to split the concept of students' learning effectiveness into three new concepts: *effectiveness of students with peers*, *effectiveness of students with instructors*, and *effectiveness of students with contents*. In online courses, such as MOOCs and SPOCs, the first and the second kinds of effectiveness can be measured by considering the number and quality of the messages submitted by students in social tools (e.g., discussion forums) and addressed to their peers or to the teachers; the third kind of effectiveness can be measured considering the number and type of educational resources and activities completed by students.

Delving a little deeper into the effectiveness of students with contents, authors in [15] collected students' low-level interactions with educational resources offered as videos (annotating interactions such as clicking "*next*" to skip the video, "*prev*" to go back, or movements over the video content) and concluded that videos that provide individual control to content (instead of random access) lead to a higher learning effectiveness. Moreover, Feng et al. [16] found that the final students' scores are correlated with specific students' actions, such as requesting for hints when solving automatic correction exercises with the support of an Intelligent Tutoring System (ITS). These authors also proposed a model to predict future scores based on students' interactions on an ITS. Therefore, these two works support Swan's thesis about the existing relationship between the effectiveness of students with contents and students' learning effectiveness.

Currently, there are several Learning Analytics (LA) techniques that enable to capture students' low-level interactions with educational resources and activities in online courses [8] and that can thus be employed to calculate the effectiveness of students with contents in MOOCs and SPOCs. Low-level events collected through these techniques can be transformed into datasets, useful to understand students' higher level behaviour [9]. For example, Blikstein [17] collected low-level interactions, such as "key presses", "button clicks" and changes in code in programming activities, categorizing students according to their performance (e.g., "*copy and pasters*", "*self-sufficient*", etc.). Other types of higher level profiles inferred from low-level interactions with educational resources and activities that have been reported in the literature for students who watched videos and solved automatic correction exercises are "*hint abuser*", "*hint avoider*", "*student misuse*", "*video avoider*", "*unreflective user*" or "*procrastinator*" [18, 19, 20, 21, 22, 23].

Despite these studies and the importance of LA in current online courses, especially in MOOCs and SPOCs, the current literature shows few works that have proposed metrics to calculate the effectiveness of students with contents in a precise manner. Actually, most e-learning platforms compute at most the number of educational resources and activities completed and the number of attempts and the grade obtained in them; but do not take into account the structure of the educational resources and activities and how they relate to each other within the course. Existing metrics mainly focus on characterizing the interactions among peers or with instructors in discussion boards [24], adapting the contents to students' goals and previous knowledge [25], and helping students advance their learning tasks through hints [26].

In this study, we adopt the main outcomes in the literature of LA and students' learning effectiveness to propose a Precise Effectiveness Strategy (PES) that supports the definition of metrics for calculating the *effectiveness of students when interacting with contents* (educational resources and activities) in online courses, including both MOOCs and SPOCs. The aim of PES is to advance the challenge of organizing the vast amounts of students' interaction data available in MOOCs and SPOCs coming from multiple and heterogeneous sources, to create useful datasets that facilitate the simple and precise representation of the effectiveness of students with educational resources and activities in a course, and the comparison of several courses using the overall effectiveness of their students.

3. THE PRECISE EFFECTIVENESS STRATEGY (PES)

The precise effectiveness strategy (PES) is a generic methodology for determining the *effectiveness of students when interacting with educational resources and activities* in online courses. This methodology includes four phases, which must be followed in for defining precise metrics for calculating this effectiveness. The different factors of the specific learning context are addressed throughout the four phases. Therefore, PES does not define metrics itself, but supports their progressive definition, as in the case study provided in Section 4.

The factors considered by PES in the progressive definition of metrics include: the kind of educational resource or activity that the student has to complete; the particular characteristics and structure of the educational resource or activity; the part/s of the educational resource or activity that the student has already completed; the relationships between the different educational resources and/or activities within the course. Regarding the latter, it is important to point out the base assumption that students achieve a higher effectiveness when completing educational resources and activities that do not repeat concepts that are present in previous resources and activities. Moreover, it is important to note that PES does not take into account the time spent to complete the educational resources and activities. Accordingly, two students would be equally effective if they manage to complete the same resources and activities in the same course. Finally, PES establishes that resource full completion implies the correct interaction with the activity (e.g. solving correctly an exercise but not just accessing and attempting it).

PES proposes a process for defining the metrics and calculating the effectiveness of students with educational resources and activities that comprises four phases.

- *Phase I: To select the educational resources and activities*. From the many educational resources and activities that can appear in an online course, those that the teacher wants to consider in the analysis of the effectiveness are selected in this phase (e.g., all the video lectures and all the parametric exercises). In this phase the teacher can select the resources and activities depending for instance on their number of occurrences and weight in the final grade, discarding those that do not cover learning contents (e.g. presentation videos).

- *Phase II: To calculate the effectiveness for each individual resource and activity*. A metric for calculating the effectiveness must be provided for each educational resource and activity. These metrics will be mathematical functions that represent continuous variables in the range [0-1]. In order to define these metrics, the common and specific characteristics of each resource and activity must be taken into account. For instance, a common characteristic in video-type resources can be that the first 15 seconds are for the institutional presentation, and thus watching them does not increase the effectiveness. In contrast, a specific characteristic can be the existence of fragments in a video that present the most important information of the course. Another specific characteristic can be the existence of fragments that are closely related to others in the same video. All these example factors (and others that are considered meaningful) must be taken into account when defining the metrics in this phase.

- *Phase III: To calculate the effectiveness for all the resources and activities of the same type*. In this phase the effectiveness of each educational resource and activity is weighted considering all the resources and activities of the same type that were selected in Phase I. That is, for each type of resource, weights are assigned to each individual resource considering its importance in the course and its relationships with other resources of the same type; resources with a lower importance and closely related to others receive lower weights. The same applies to each individual activity. At the end of this phase a metric will be provided for each type of resource and for each type of activity represented as a mathematical function with a continuous variable in the range [0-1].
- *Phase IV: To calculate the global effectiveness of the course*. In the last phase, the effectiveness of each educational resource and activity is weighted considering all the resources and activities selected in Phase I. The factors to be taken into account in this phase include the relationships among different activities and resources (e.g., if the student must watch a video before solving an exercise). At the end of this phase a global metric will be provided as a mathematical function with a continuous variable in the range [0-1].

4. CASE STUDY

PES has already been applied in the context of zero courses for freshmen at a Spanish University. Zero courses are remedial courses for students that register for a first year degree but need to review basic concepts of mathematics, physics or chemistry. As an additional support for these courses, the University provides an online environment with videos and exercises so that students can practice thoroughly the different topics. These online environments represent a clear example of SPOCs.

Two editions of these zero courses have been carried out so far. The first edition (2012) included a SPOC in physics, while in the second edition (2013) there were three SPOCs in physics, mathematics and chemistry. The Khan Academy platform was chosen as the online environment for the SPOCs deployment. This platform supports the types of contents required for these courses: video lectures and automatic correction exercises.

Teachers uploaded about 30 video lectures per SPOC. There was at least one exercise related to each video. Most of the exercises were parametric (i.e. after completing an exercise, students could repeat it again, its variables receiving new values), although there were also a few multiple choice exercises. Students could solve correctly each parametric exercise several times to gain mastery. Students could also solve correctly several multiple choice exercises of the same type addressing similar topics.

4.1. PHASE I: SELECTION OF EDUCATIONAL RESOURCES AND ACTIVITIES

The PES methodology was employed in this case study to define the metrics that enabled to calculate the effectiveness of students with the two kinds of educational resources and activities available in these SPOCs: videos and exercises. In Phase I, all the videos and exercises uploaded in each SPOC by teachers were selected. That included: 27 videos and

35 exercises in physics (2012 edition), 30 videos and 30 exercises in physics (2013 edition), 25 videos and 30 exercises in mathematics, and 22 videos and 49 exercises in chemistry.

4.2. PHASE II: CALCULATION OF THE EFFECTIVENESS FOR EACH INDIVIDUAL RESOURCE AND ACTIVITY

In Phase II, metrics for calculating the effectiveness of students with each individual video and exercise were defined. For this calculation, a set of assumptions regarding the characteristics of the videos and exercises were made (subsection 4.2.1). Based on these characteristics, some initial hypotheses were proposed in order to define the metrics (subsection 4.2.2). These hypotheses were validated by experts, who also contributed in the definition of the mathematical functions for these metrics (subsection 4.2.3).

4.2.1. CHARACTERISTICS OF THE VIDEOS AND EXERCISES

The analysis of the video lectures revealed that they all followed an equivalent structure.

- The length of the video was short (about 10 minutes). Thus, it was expected that students could maintain the attention during all the time, so the effectiveness should not be different because of a change of students' attention when interacting with the resource.

- All the parts in the video were equally important in general.

- Different parts in the video were connected to other parts of the same video in a way that if a student missed some part, he could not understand the whole concept. So, a student that watched only one part was not proportionally effective with that video.

Regarding exercises, the analysis revealed that they all followed an equivalent structure too.

- The first time that a student solved an exercise correctly entailed a higher level of difficulty than the following times the students attempted the same exercise (in parametric ones) or a related exercise (in multiple choice ones). This is because in each parametric exercise the statement was always the same but the values of the variables changed when the exercise was solved correctly, while each multiple choice exercise was related to previous ones that were solved correctly.

4.2.2. INITIAL HYPOTHESES

Based on the aforementioned characteristics the following initial hypotheses were formulated.

- H1: The effectiveness of a student with a video is 1 if the student has watched 100% of the video and 0 if the student has not started watching the video.
- H2: If a student has watched the first half of a video, he has achieved less than half the effectiveness that can be potentially achieved with that video. This can be generalized so if a student has watched some part of a video, he has achieved less than the proportional effectiveness for that part. This is due to the existing

connections between different parts in a video, so watching the whole video gives an effectiveness which is greater than the sum of the effectiveness of watching just the different parts.

- H3: The effectiveness of a student with an exercise is 1 if the student has solved it correctly N times (with N a variable that might depend on the exercise) and 0 if the student has never solved it correctly.
- H4: If a student has solved an exercise correctly M times (with M<N), then the probability of making the exercise correctly again is greater than it was when the student had solved (or attempted to solve) it M-1 times.

4.2.3. VALIDATION BY EXPERTS

The use of experts is a widespread technique in research on educational technologies [27], [28]. A total of 8 educators and learning analytics experts, which had no relation with this research, participated in the validation of the previous hypotheses (1 PhD Visiting Associate Professor, 2 Teaching Assistants, 2 Research Assistants, and 3 undergraduate students working on learning analytics topics in their final Master's Thesis). In addition, these participants contributed to the definition of the mathematical functions that enabled to calculate the effectiveness of students with videos and exercises in this case study.

First, experts watched two selected representative videos and interacted with three selected representative exercises extracted from the four SPOCs, so they could figure out their characteristics. Next, they were asked: a) to fill in a survey assessing the four abovementioned hypotheses; b) to draw a mathematical function representing the effectiveness in the case of videos and the effectiveness in the case of exercises; and c) to write free comments in an open question.

Experts were asked directly about their level of agreement with hypotheses H2 and H4 in a scale from 1 (completely disagree) to 5 (completely agree). Participants highly agree with hypothesis H4 (6 experts rated with a 5 and 2 with a 4), while for H2 there was a moderate agreement, since 2 experts rated this hypothesis with a 5, 3 with a 4 and 3 with a 3.

As experts had to draw a function representing the video effectiveness (Y axis) versus the percentage of completed video (X axis), and another function representing the exercise effectiveness (Y axis) versus the number of correct exercises of the same type (X axis), then the hypotheses H1 and H3 were evaluated by experts. There was a complete agreement with hypotheses H1 and H3 from all 8 experts, so the hypotheses were validated. In the case of hypothesis H3, 6 of the experts indicated N=8 as a suitable value (so these experts considered that students should solve 8 exercises correctly of the same type to obtain an effectiveness of 1 in that resource), while the other two experts proposed N=7 and N=5.

Previously to the experts' evaluation, two of the authors of this paper carefully analyzed the videos and exercises of the four SPOCs. Based on their analysis, they drew the effectiveness functions for videos and exercises. Next, as commented, the 8 experts also had to draw these functions in order to validate the functions provided by the authors of the paper.

Figure 1 represents the effectiveness of students with a video (axis y) depending on the percentage of video completed by this student (axis x). Line b is the proposed by the authors, while line d is a curve calculated as the mean of the curves proposed by the 8 experts. It is noteworthy that this is a not a linear function due to the aforementioned characteristics of videos and the hypotheses validated by the experts. For a higher precision, a polynomial that represents the curve b was defined as follows:

$$Effectiveness_{Video} = 0.003 + 0.3816x + 0.1793x^2 + 0.0998x^3 + 0.0206x^4 + 0.0185x^5$$

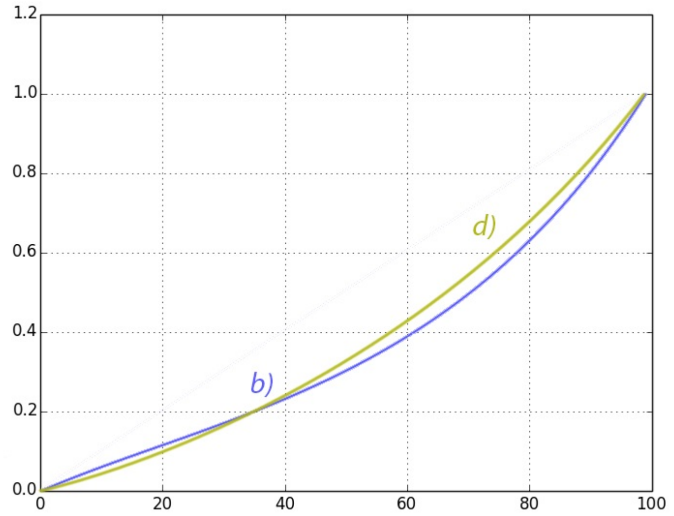

Figure 1. Polynomial representation to calculate the effectiveness of students with a video in the case study.

Figure 2 represents the effectiveness of students with a parametric exercise (axis y) depending on the number of times the student solved correctly that exercise (with N=8 where the effectiveness is 1). Line b is the proposed by the authors, while line d is a curve calculated as the mean of the curves proposed by the 8 experts. As in the case of videos, this is not a linear function, although in this case, the increase of effectiveness decreases as the student correctly solves the same parametric exercise several times. For a higher precision, a polynomial that represent curve b is defined as follows:

$$Effectiveness_{Exercise} = 0.65 + 42.85x - 8.33x^2 + 9.04x^3 - 6.66x^4 - 1.9x^5$$

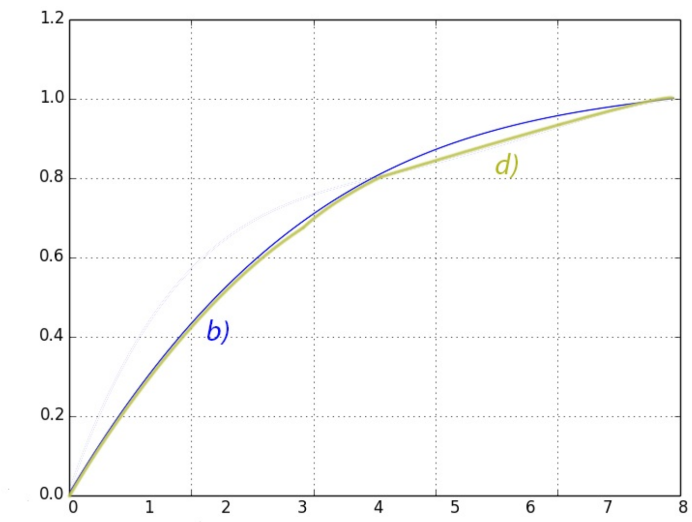

Figure 2. Polynomial representation to calculate the effectiveness of students with a parametric exercise in the case study.

As the curves proposed by the authors were both (for videos and exercises) very close to the ones calculated as the means of the curves by experts, then the former curves were validated. We took the curves proposed by the authors as base for the calculation of effectiveness of this research, as these were a result of a more careful analysis of videos and exercises for the 4 SPOCs. If the curves had been quite different, then we would have analyzed the reasons for it to make the proper changes. In any case, calculations might be easily described with the curves of experts, although there would not be significant changes, as they are very close.

4.3. PHASE III: CALCULATION OF THE EFFECTIVENESS FOR ALL THE RESOURCES AND ACTIVITIES OF THE SAME TYPE

Next, in Phase III, we calculated the effectiveness for all the videos and the effectiveness for all the exercises. In order to simplify the mathematical functions, we assumed that the content in each video was independent of that in other videos, which was true in most of the videos. Therefore, we considered that the effectiveness of students with videos was a weighted average of the effectiveness of students with each video. The same approach was followed in the case of exercises. Both equations are presented next.

$$Effectiveness_{Videos} = \frac{1}{N} * \sum_{1}^{N} Effectiveness_{Video\_i}$$

$$Effectiveness_{Exercises} = \frac{1}{N} * \sum_{1}^{N} Effectiveness_{Exercise\_i}$$

4.4. PHASE IV: CALCULATION OF THE GLOBAL EFFECTIVENESS OF THE COURSE

Finally, to provide a metric for the global effectiveness of each SPOC in this case study (Phase IV), we calculated the weighted average of the effectiveness of students with videos and the effectiveness of students with exercises as indicated (see the equation below). The equation has been simplified for illustrative purposes assuming that each video had one and only one associated exercise, although some SPOCs had a few more exercises than videos (see section 4.1).

$$Effectiveness = \frac{Effectiveness_{Videos} + Effectiveness_{Exercises}}{2}$$

4.5. Discussion about the generalization of the metrics proposed in this case study

The metrics presented in this section were obtained following the PES methodology. These concrete metrics can be more or less precise depending on the characteristics of the educational context and the assumptions made. For example, the importance of the different parts within a video influences the metric that supports the calculation of the effectiveness of students with that video (Phase II in PES). Also, the existence of several similar exercises to reinforce a concept influences the metric for the calculation of the effectiveness of students with exercises (Phase III in PES). In the case study, although a few videos and exercises had some particularities, they were in general quite homogeneous because teachers were required to follow a set of given rules. This fact supported the starting assumptions of applying the same function to all the videos and the same function to all the exercises in Phase II.

A more precise analysis could have been done to define different metrics for each video and/or for each exercise in the case study instead of a generalization for all of them. For example, some parametric exercises are easier to solve them correctly once they have been solved correctly before, so they can have a different function for the calculation of the effectiveness. Nevertheless, there is a trade-off between the precision in the calculation of the effectiveness and the complexity of the metrics and processes. In this case study, we consider that the proposed metrics are precise enough, although they could be improved adding more complexity to them.

The defined metrics of the case study cannot be reused for all the educational contexts. For each educational context, the PES methodology should be followed to define other metrics. In this direction, the experts provided different comments in the free open question of the survey, such as “As a difference to the parametric exercises, the explanation of a concept in a video can differ a lot depending on the concept, the culture, or the methodology used, so it is difficult to generalize” or “For the case of videos, the function can vary in areas different from engineering which can be more theoretical. In these areas, the effectiveness can be more lineal”.

5. Results and discussion

This section presents and discusses the effectiveness of students with videos and exercises in the case study, considering the metrics defined in the previous section. After that, a comparison of the global effectiveness between the four SPOCs is carried out; that comparison aims to help teachers make quick and informed decisions about whether videos and/or exercises need to be rethought in future editions of the SPOC. Finally, the metrics defined in the previous section are related to others found in the literature for calculating behavior patterns that aims to help teachers make decisions about whether to promote or to avoid certain behavior patterns in order to achieve a higher level of effectiveness in their courses.

5.1. Analysis of the Effectiveness

The metrics defined using PES enable teachers and other educational stakeholders to have evidences of students' progress and interactions with the educational resources and activities in a MOOC or SPOC. These metrics can be used and represented in very different ways:

- Teachers can easily view the interactions of students with educational resources and activities.
- Students can be classified in clusters depending on their effectiveness with videos, with exercises or with a combination of both.
- Teachers can set thresholds so that if students do not achieve a given effectiveness level, they will not pass the course.

Visualizations representing the class effectiveness using the metrics defined through PES are particularly useful in massive groups of students. As an example, Figure 3 represents the effectiveness of each student with videos and exercises in the SPOC in mathematics; similar graphs can be obtained for the SPOCs in physics and chemistry. Each point represents the effectiveness of a particular student with videos (axis Y) and exercises (axis X). This way, teachers can know students' interactions with educational resources and activities, and observe at a glance different clusters:

- Low effectiveness with videos and exercises.
- Medium effectiveness with videos and exercises.
- High effectiveness with videos and exercises.
- High effectiveness with videos, but low effectiveness with exercises and vice versa.

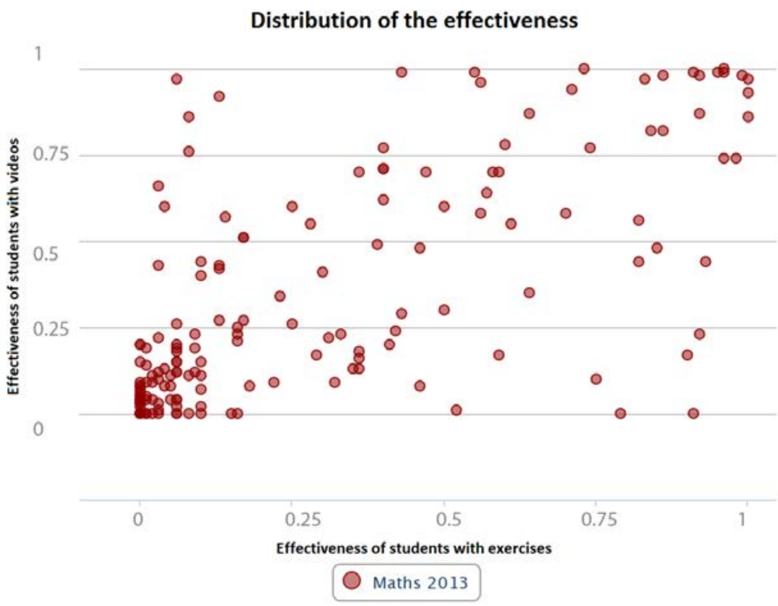

Figure 3. Analysis of the effectiveness of students with videos and exercises in the SPOC in maths.

Figure 4 classifies the students in Figure 3 in 5 clusters regarding their level of effectiveness with videos and exercises. This visualization is especially useful to detect those participants who interact little with contents (high proportion of students in area tagged “1”), or those resources and activities that are poorly balanced (high proportion of students in areas tagged “2” and “3”).

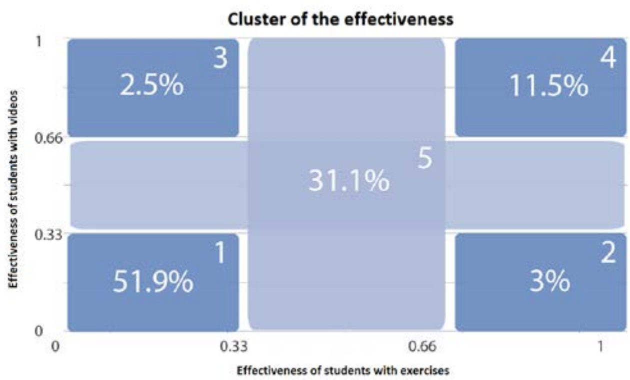

Figure 4. Clusters representing the effectiveness of students with videos and exercises in the SPOC in maths.

In addition, Figure 5 represents the effectiveness of each student in the four SPOCs used in the case study. This representation facilitates a quick comparison of the effectiveness of students with videos and exercises among courses. Interesting conclusions can be drawn from this graph, such as in which course was the dropout rate higher or in which course students achieved a higher effectiveness with videos or with exercises.

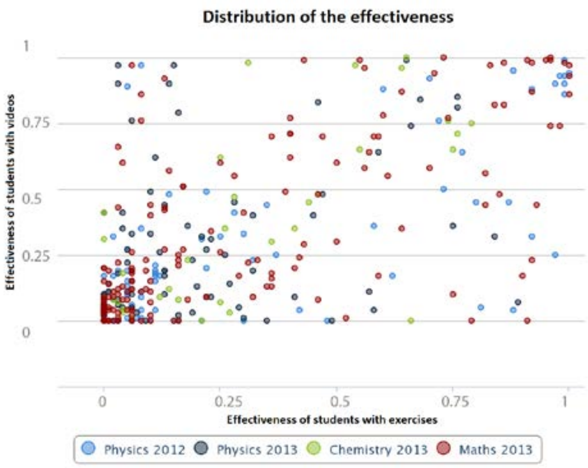

Figure 5. Analysis of the effectiveness of students with videos and exercises in the four SPOCs.

5.2. COMPARISON OF THE EFFECTIVENESS AMONG COURSES

The metrics defined through PES can also be useful to compare courses according to the global effectiveness of the students with educational resources and activities. To do so, the mean of the effectiveness of each student can be calculated and represented as in Figure 6, which compares the four SPOCs under study. It is noteworthy that each point in Figure 6 represents a course, and that the size of the point is proportional to the number of students enrolled in that course. This representation reveals courses with a higher dropout rate so that their teachers can complement lectures with additional resources, or offer hints that help students resolve the exercises. This representation also serves to detect courses with misalignments between videos and exercises, so that teachers can rethink the videos and exercises using as models those courses with a higher global effectiveness of students.

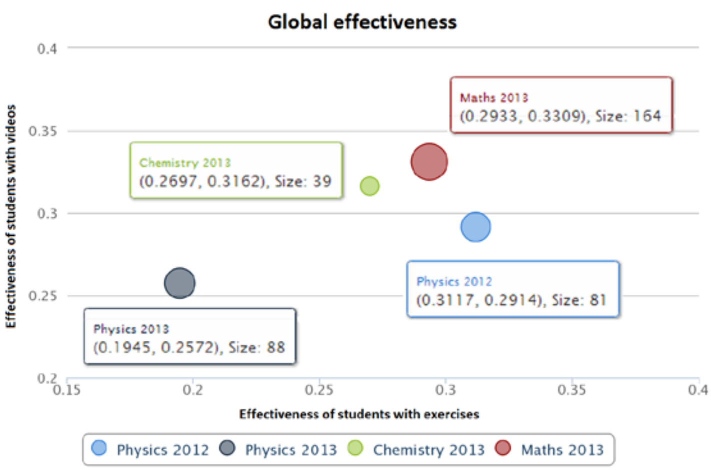

Figure 6. Comparison of the effectiveness of students with videos and exercises in the four SPOCs.

Although graphical comparisons are very useful, statistical comparisons can also help determine if there are statistically significant differences between courses. Tables 1 and 2 present the four SPOCs under study indicating the number of students, and the mean and standard deviation for the effectiveness with videos and with exercises. Students enrolled in several courses were only taken into account once, being randomly assigned to one of the courses with the condition that each course received the same number of these students (so that samples are independent).

A one-way between subjects ANOVA was carried out for the effectiveness of students with videos, taking the type of course as factor, which gives a result of ($F=1.085$, $p=0.356$). Based on the significance related to the type of course, we can conclude that the type of course did not have a considerable effect on the effectiveness of students with videos.

A one-way between subjects ANOVA was carried out for the effectiveness of students with exercises, taking the type of course as factor, which gives a result of ($F=2.597$, $p=0.052$). In this case, the type of course is statistically significant at the 90% level. As the number of samples is high but quite different in each course and we cannot assume homogeneity of variances, then the Games-Howell test was selected as post-hoc test to compare all the courses in pairs. In two cases statistically significant differences were found at the 90% level: between physics 2012 and physics 2013 in favor of physics 2012 ($p=0.059$); and between maths 2013 and physics 2013 in favor of maths 2013 ($p = 0.029$). This result confirms that effectiveness of students with exercises was lower in physics 2013 than in physics 2012 and maths 2013.

	# students	**Mean**	**Std. dev.**	**Conf. interval (95%)**
Physics 2012	81	0.29	0.32	[0.22, 0.36]
Physics 2013	88	0.26	0.29	[0.20, 0.32]
Maths 2013	164	0.33	0.33	[0.28, 0.38]
Chemistry 2013	39	0.32	0.33	[0.21, 0.42]

Table 1: Effectiveness of students with videos per course.

	# students	**Mean**	**Std. dev.**	**Conf. interval (95%)**
Physics 2012	81	0.31	0.35	[0.23, 0.39]
Physics 2013	88	0.19	0.23	[0.15, 0.24]
Maths 2013	164	0.29	0.33	[0.24, 0.34]
Chemistry 2013	39	0.27	0.26	[0.18, 0.36]

Table 2: Effectiveness of students with exercises per course.

5.3. RELATIONSHIPS AMONG METRICS

The metrics defined through PES can also be related to other higher level metrics published in the literature defining students' behavior patterns when watching videos and solving exercises [18]. These metrics are also in the range [0-1] and are:

- *Recommendation listener*. This metric gives an idea if a student follows the system recommendations or not about the next exercises to solve.

- *Hint avoider*. This metric shows if a student avoids selecting hints when he should request them for solving an exercise.

- *Hint abuser*. This metric indicates if a student requests for hints when he does not have to for solving an exercise.

- *Video avoider*. This metric gives information about if a student avoids watching videos when he should watch them.

- *Unreflective student*. This metric gives information about if a student answers too quickly to an exercise when he should reflect more.

Table 3 shows the Pearson Correlation (N=372, two-tailed significance) among the effectiveness of students with videos and with exercises (as calculated in the case study) and the aforementioned behavior metrics. Significance values at the 95% level are marked with an asterisk.

	Effectiveness of students with videos	**Effectiveness of students with exercises**
Effectiveness of students with videos	1	0.640 (Sig. 0.000)
Effectiveness of students with exercises	0.640* (p= 0.000)	1
Recommendation listener	-0.130* (p= 0.012)	-0.137* (p= 0.008)
Hint avoider	0.107* (p= 0.039)	0.032 (p= 0.544)
Hint abuser	-0.092 (sig. 0.077)	-0.134* (p= 0.010)
Video avoider	-0.228* (p= 0.000)	- 0.107* (p= 0.039)
Unreflective student	0.048 (p= 0.353)	0.003 (p= 0.961)

Table 3. Pearson correlation among the effectiveness of students with videos and exercises and other students' behavior metrics found in the literature.

As expected, the data shows that the effectiveness of students with videos is strongly correlated with the effectiveness of students with exercises. This correlation indicates that active users, who interacts a lot with one of these types of contents, he will potentially interact much also with the remaining types of contents. There are also several statistically significant relationships between the effectiveness of students with videos and exercises and other students' behavior metrics at 95% level. In all the cases, correlations are moderated and in most cases with a negative correlation:

- **Recommendation listener and effectiveness of students with videos and with exercises.** This relationship indicates that students who followed the recommendations provided by the Khan Academy platform are less effective when interacting with videos and exercises.

- **Video avoider and effectiveness of students with videos and with exercises.** This relationship indicates that students who avoid watching videos are less effective when interacting with the videos they access (as one could expect), and also when interacting with exercises, particularly those related with the videos that the student avoided watching.

- **Hint abuse and effectiveness of students with exercises.** This relationship indicates that abusing of hints has a negative impact on the effectiveness of students with exercises. This result contradicts the idea that abusing of hints entails solving a higher number of exercises correctly and quicker.
- **Hint avoider and effectiveness of students with videos.** This relationship indicates that students that avoid hints are more effective when interacting with videos.

6. Conclusions and Future Work

This paper presents the Precise Effectiveness Strategy (PES), which is a generic methodology for defining metrics to calculate the effectiveness of students with educational resources and activities in MOOCs and SPOCs. PES establishes a set of four phases along which these metrics are refined, including the particular characteristics of the educational resources and activities, and the educational context. Although PES can be applied to online courses in general, it is especially meaningful in MOOCs, since they involve many students, and teachers cannot devote a lot of time in analyzing students' low-level interactions with contents, but need automatic and precise metrics to know and visualize the effectiveness of students with educational resources and activities.

PES has been applied to four SPOCs based on video lectures and automatic correction exercises, and deployed in the Khan Academy platform. Specific metrics based on non-linear functions were proposed for calculating the effectiveness of students with videos and exercises in these courses. Several visualizations of these effectiveness metrics were presented to help teachers understand students' interactions with contents and help them make informed decisions about the videos and exercises they included in these courses. The application of statistical techniques discovered some statistically significant differences between the effectiveness of students with videos and exercises in some of the SPOCs under study. This type of information can be useful for the institution in which the SPOCs ran. In addition, some interesting relationships among the effectiveness metrics defined in these case study and other behavior metrics published in the literature were found.

PES is a first effort towards a more precise calculation of the effectiveness in terms of students' interaction with educational resources and activities in online learning environments. However, this first approach raises several questions and aspects that should be discussed and pursued in future work. The first aspect is to what extend the metrics defined in this case study serve for characterizing students' interactions with contents in other MOOCs and SPOCs deployed in platforms other than the Khan Academy. This paper presented a case study with four SPOCs deployed in the Kahn Academy, but other courses deployed in the edX platform are planned for the near future in order to understand the applicability of the metrics defined here. Another line of work is the analysis of the results obtained in the case study in relation to students' knowledge acquisition. New experiments including pre and post-tests to analyze students' learning gains will help on re-shaping these metrics for better capturing the concept of effectiveness. In addition, an authoring system might be incorporated so that the creators of educational resources and activities (e.g. videos and exercises) might draw the different effectiveness functions.

All in all, this paper aims to reflect on the importance of defining metrics able of extracting relevant high-level information about the effectiveness of students in MOOCs and SPOCs from raw data captured by the online platform. We contend that both PES and the metrics defined in the case study deserve further research in order to understand their implications as a support for teachers and institutions in online courses.

REFERENCES

1. Pappano, L. The year of the MOOC. (2012), http://www.nytimes.com/2012/11/04/education/edlife/massive-open-online-courses-are-multiplying-at-a-rapid-pace.html

2. Belenger, Y., Thornton, J., Bioelectricity: A Quantitative Approach, Duke University's First MOOC. Duke University Libraries (2013), 1-21.

3. Kolowich, S. The MOOC 'Revolution' May not Be as Disruptive as Some Had Imagined, The Chronicle of Higher Education (2013), http://chronicle.com/article/MOOCs-May-Not-Be-So-Disruptive/140965

4. Tucker, B. The Flipped Classroom: Online instruction at home frees class time for learning, *Education Next*, (Winter 2012), 82-83.

5. Coughlan, S. Harvard plans to boldly go with 'Spocs', BBC News Business (2013), http://www.bbc.co.uk/news/business-24166247

6. Price, G. Are We Beginning the Post-MOOC ERA? Two SPOCs (Small Private Online Courses) are Underway From Harvard/edX (2013), http://www.infodocket.com/2013/09/27/are-we-already-in-the-post-mooc-era-two-spocs-small-private-online-courses-are-underway-at-harvardedx

7. Siemens, G., Long, P. Penetrating the Fog: Analytics in Learning and Education, *Educause Review*, September/October (2011), 31-40.

8. Siemens, G. Learning analytics: envisioning a research discipline and a domain of practice. In *Proc. LAK' 12*, (2012), 4-8.

9. Clow, D. The learning analytics cycle: closing the loop effectively, In *Proc. LAK' 12*, (2012), 134-138.

10. Hiltz, S.R., Arbaugh, J.B. Improving Quantitative Research Methods in studies of Asynchronous learning Networks (ALN), In J. Bourne & J. C. Moore (Eds.) *Elements of Quality Online Education, Practice and Direction*. Needham, MA: Sloan Center for Online Education, (2003), 59-72.

11. Swan, K. Learning effectiveness: what the research tells us. In J. Bourne & J. C. Moore (Eds.) *Elements of Quality Online Education, Practice and Direction.* Needham, MA: Sloan Center for Online Education, (2003), 13-45.

12. Ni, A.Y. Comparing the Effectiveness of Classroom and Online Learning: Teaching Research Methods, *Journal of Public Affairs Education*, 19, 2 (2013), 199-215.

13. Moody, D.L., Sindre, G. Evaluating the Effectiveness of Learning Interventions: An Information Systems Case Study, In *Proc. ECIS 2003*, (2003), 1311-1326.

14.Rourke, L., Anderson, T., Garrison, D. R., Archer, W. Assessing social presence in asynchronous text-based computer conferencing, *Journal of Distance Education*, 14 (2), 50-71, 1999.

15.Zhang, D., Zhou, L., Briggs, R.O., Nunamaker Jr., J.F. Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness, *Information & Management*, 43, 1 (2006), 15-27.

16.Feng, M., Heffernan, N.T., Koedinger, K.R. Predicting state test scores better with intelligent tutoring systems: developing metrics to measure assistance required. In *Proc. ITS 2006,* (2006), 31–40.

17.Blikstein, P. Using learning analytics to assess students' behavior in open-ended programming tasks. In *Proc. LAK' 11*, (2011), 110-116.

18.Muñoz-Merino, P.J., Ruipérez-Valiente, J.A., Delgado-Kloos, C. Inferring higher level learning information from low level data for the Khan Academy platform, In Proc. LAK' 13, (2013), 112-116.

19.Aleven, V., McLaren, B.M., Roll, I., Koedinger, K.R. Toward Tutoring Help Seeking: Applying Cognitive Modeling to Meta-cognitive Skills. In *Proc. ITS 2004*, (2004), 227–239.

20.Aleven, V., McLaren, B.M., Roll, I., Koedinger, K.R. Toward Meta-cognitive Tutoring: A Model of Help Seeking with a Cognitive Tutor, *Int. Journal of Artificial Intelligence in Education*, 16, 2 (2006), 101–128.

21.Baker, R.S., Corbett, A.T., Koedinger, K.R., Wagner, A.Z. Off-Task Behavior in the Cognitive Tutor Classroom: When Students "Game The System". In *Proc. CHI 2004* (2004), 383-390.

22.Baker, R.S., Corbett, A.T., Koedinger, K.R. Detecting Student Misuse of Intelligent Tutoring Systems. In *Proc. ITS 2004*, (2004), 531-540.

23.Tervakari, A-M., Marttila, J., Kailanto, M., Huhtamäki, J., Koro, J., Silius, K. Developing Learning Analytics for TUT Circle. In Proc. IFIP AICT 2013, (2013), 101-110

24.Bliss, C.A., Lawrence, B. From posts to patterns: A metric to characterize discussion board activity in online courses. *Journal of Asynchronous Learning Networks*, 13, 2 (2009), 15-32.

25.Brusilovsky, P., Vassileva, J. Course sequencing techniques for large-scale web-based education. *Int. Journal of Continuing Engineering Education and Life Long Learning*, 13, 1 (2003), 75-94.

26.Muñoz Merino, P. J., Delgado Kloos, C. A software player for providing hints in problem based learning according to a new specification. Computer Applications in Engineering Education, 17, 3 (2009), 272-284

27.Pardos, Z.A., Baker, R., San Pedro, M., Gowda, S.M., Gowda, S.M. Affective states and state tests: investigating how affect throughout the school year predicts end of year learning outcomes. In Proc. *LAK 2013(2013).*

28.*Suebnukarn, S., Haddawy, P.* A Bayesian approach to generating tutorial hints in a collaborative medical problem-based learning system. Artif. Intell. Med, 38 (2006), 5-24.