Muñoz-Merino, P. J., Ruipérez-Valiente, J. A., Alario-Hoyos, C., Pérez-Sanagustín, M., & Delgado Kloos, C. (2015). Precise effectiveness strategy for analyzing the effectiveness of students with educational resources and activities in MOOCs. Computers in Human Behavior, 47, 108-118.

doi: https://doi.org/10.1016/j.chb.2014.10.003

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


## 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 and illustrates the application of the Precise Effectiveness Strategy (PES). PES is a generic methodology for defining precise metrics that enable 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 visualizations within and between courses 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 behavioral patterns. Results from using PES in the case study revealed that the effectiveness metrics proposed had a moderate negative correlation with some behavioral patterns like recommendation 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 (Pappano, 2012). 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 (Belenger & Thornton, 2013). These types of MOOCs, also known as xMOOCs, are the most widespread, and follow the so-called broadcast model (Kolowich, 2013).

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 these contexts, MOOCs are exploited to enhance teaching and learning in the form of e.g., 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) (Tucker, 2012). 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 (Coughlan, 2013). 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 (Price, 2013).

In both, MOOCs and SPOCs, there can be a very large number of students, so teachers need precise strategies to know what is going on with each individual student and with the whole class. Current learning analytics techniques (Siemens & Long, 2011) enable the collection of huge amounts of low-level data regarding for instance students' interaction among themselves or with educational resources and activities, both in MOOCs and SPOCs (Siemens, 2012). 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 (Clow, 2012).

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, there is a need of 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 different 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., the number of videos completed, the number of exercises accessed or the number of exercises correctly solved). Furthermore, 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 the effectiveness of thousands of students at a glance 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 students' progress in a video.

In this context, the specific research questions addressed in this work are:

- Can we propose a general strategy for calculating the effectiveness of students when interacting with resources and activities in a precise way, giving some general guidelines on how to apply it?
- How can we particularize the general strategy for calculating the effectiveness of students with resources and activities in a case study with an intensive use of videos and exercises (which are some of the most commonly-used contents in MOOCs and SPOCs)?
- How can we use the effectiveness metrics to compare students, contents and courses?
- Are the effectiveness metrics related to other important student's behavioral patterns?

This paper presents Precise Effectiveness Strategy (PES), a generic methodology for supporting the calculation of the effectiveness of students when interacting with educational resources and activities. PES guides the process for defining metrics to calculate this effectiveness, considering the main particularities of the learning context (e.g., the relationships among the different video resources). 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 enable: 1) the representation of the effectiveness of students' interactions with videos and exercises in a simple way; 2) the measurement of the relationship of the effectiveness metrics with behavioral patterns; and 3) the comparison of 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 learn, how well did they master skills and how well can they apply knowledge" (Hiltz & Arbaugh, 2003). 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 (Swan, 2003; Ni, 2013). In order to measure learning effectiveness, most authors typically use (if possible) achievement tests or surveys for collecting student perceptions (Moody & Sindre, 2003).

The study by Swan (2003) 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) (Rourke, Anderson, Garrison & Archer, 1999). 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 type (e.g., question, answer, etc.) of messages submitted by students in discussion forums and addressed to their peers or to the teachers, as well as their quality (e.g., measured through voting systems or Natural Language Processing approaches); 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, Zhang, Zhou, Briggs and Nunamaker Jr. (2006) 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, Heffernan and Koedinger (2006) found that the final students' scores are correlated with specific students' actions, such as asking 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 techniques that enable to capture students' low-level interactions with educational resources and activities in online courses (Siemens, 2012) 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 (Clow, 2012). For example, Blikstein (2011) 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*" (Muñoz-Merino, Ruipérez-Valiente, & Delgado-Kloos, 2013; Aleven, McLaren, Roll & Koedinger, 2004, 2006; Baker, Corbett, Koedinger & Wagner, 2004; Baker, Corbett & Koedinger, 2004; Tervakari, Marttila, Kailanto, Huhtamäki, Koro & Silius, 2013).

Despite these studies and the importance of learning analytics 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 (e.g., edX, Coursera, etc.) compute at most the number of educational resources and activities completed (or pending) and the number of attempts and the grade obtained in them. Some works go a step further indicating the degree of completeness of a learning content as a measure of effectiveness, but assuming that all its parts are equally important. Therefore, there is a need for more precise strategies for measuring the effectiveness of students with contents that take into account the structure of the educational resources and activities and how they relate to each other within the course. There are already a few works that address the characterization of the interactions among peers or with instructors in discussion boards (Bliss & Lawrence, 2009), adapting the contents to students' goals and previous knowledge (Brusilovsky & Vassileva, 2003), and helping students advance their learning tasks through hints (Muñoz-Merino & Delgado Kloos, 2009).

In this study, we adopt the main outcomes of these works and the literature of learning analytics to propose a Precise Effectiveness Strategy (PES). PES 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. In this context, the definition of *effectiveness of students with contents* only addresses the interactions of students with contents and not if these interactions bring learning effects or not (which is out of the scope of this paper). 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 an orderly fashion, for defining precise metrics aimed at calculating this effectiveness. The different factors of the specific learning context are addressed throughout the four phases. PES gives general guidelines, but does not define metrics itself, nor gives predetermined patterns. Instead, PES supports the progressive definition of metrics in different contexts, as in the case study provided in Section 4. The calculation of the effectiveness in a specific learning context is not a trivial task, and a detailed analysis should be done by those applying PES in order to obtain the specific metrics in each case study. It is noteworthy that PES does not focus on *learning effectiveness,* but on the *effectiveness of students with educational resources and activities*. A reasonable hypothesis is that if students are effective with educational resources and activities, then they will also be effective in learning. But the validation of this hypothesis is out of the scope of PES and of this work.

The factors considered by PES in the progressive definition of metrics include: the kinds of educational resources or activities 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 (e.g., a student could speed up or down a video or replay certain fragments). Considering time would lead to the *efficiency* of the student with contents, and would be a different problem to tackle. 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 the completion of the resource implies the correct interaction with the activity. That is, a resource is completed when a learner solves an exercise correctly, but not when he or she attempts to do it without success.

PES proposes a process for defining the metrics and calculating the effectiveness of students with educational resources and activities that comprises four phases. Figure 1 is a diagram in the Unified Modeling Language (UML) to describe PES, showing its different phases.

- *Phase I: Selection of 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, and discard those that do not cover learning contents (e.g., presentation videos). The activities should be selected in a way that all the students can obtain the maximum effectiveness regardless of the actions of other students.

- *Phase II: Calculation of 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]. The reason to work with continuous variables instead of with discrete ones is because using continuous information increases the precision in the calculation of the effectiveness. 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: Calculation of 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: Calculation of 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].

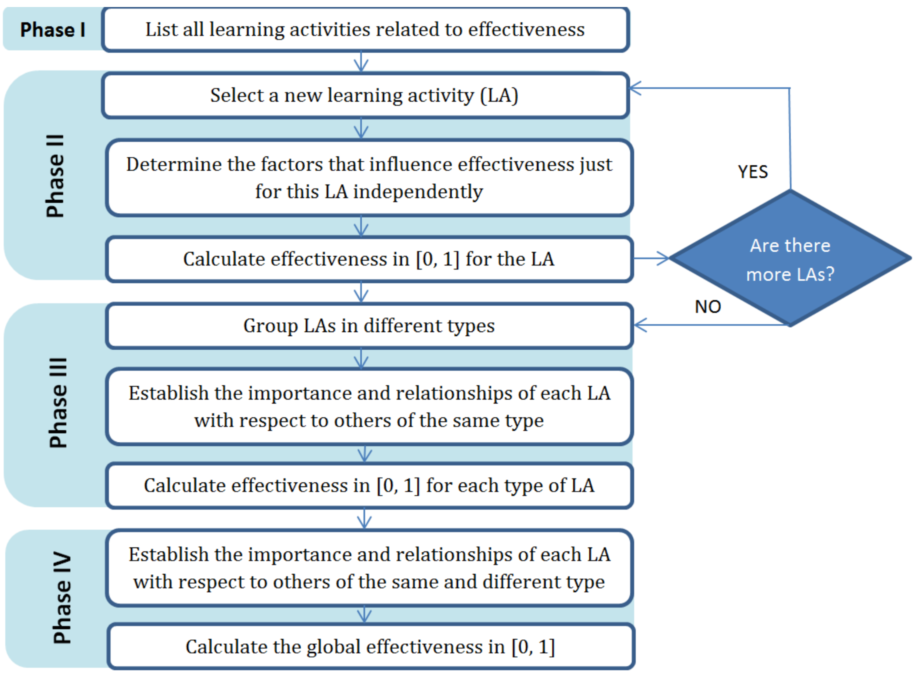


**Figure 1. A UML flow diagram describing PES**

# 4. CASE STUDY

In this section, we detail the application of PES in a particular SPOC-like learning context. . MOOCs and SPOCs typically present an intensive use of videos and exercises. Therefore, these are the two types of resources and activities that are deeply covered in this case study. This way, it is expected that the ideas that emerge from this case study can also be applied to other MOOCs and SPOCs whose videos and exercises follow similar patterns. However, the specificities of each MOOC and SPOC should be considered when applying PES. For example, the relationships among the educational resources and activities might be different, or social tools might also be included in the analysis.

PES has been applied in the context of zero courses for freshmen at a Spanish University. Zero courses are remedial courses for students who 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 SPOCs were deployed in the Khan Academy platform. 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 (but this number depends on each course). 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 enable calculation of the effectiveness of students with videos and exercises in the SPOCs. 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, the 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 followed an equivalent structure in general.

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

- All the parts in the video were equally important.

- 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 also followed an equivalent structure in general.

- 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. These hypotheses are validated by experts in the next subsection (4.2.3).

- **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 that 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 solving 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*

In order to validate the hypotheses, we invited a set of experts. Expert validation is a widespread technique in the field of Technology Enhanced Learning to get different perspectives from the results, since results are usually closely linked to the context of application (Pardos, Baker, San Pedro, Gowda & Gowda, 2013; Suebnukarn & Haddawy, 2006). A total of 8 educators and learning analytics experts external to 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). These participants contributed to the definition of the mathematical functions that enabled calculation of 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 (six experts rated with a 5 and two with a 4), while for H2 there was a moderate agreement, since two experts rated this hypothesis with a 5, three with a 4 and three with a 3. As most of the experts (five out of the eight) agreed or strongly agreed with H2, and none of them disagreed with this hypothesis (three out of eight were neutral), then we consider that H2 is validated. This is confirmed with the curve experts had to draw (see Figure 2).

In order to validate hypotheses H1 and H3, experts were asked 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). 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, six of the experts indicated N=8 as a suitable value (so these experts considered that students should solve eight exercises of the same type correctly 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, and as commented, the 8 experts also had to draw these functions in order to validate the functions provided by the authors of the paper. The experts did not know about the curves proposed by the authors and they watched two representative videos and interacted with three typical exercises for about 30 minutes. From this interaction and visualization they figured out the importance and relationship of the different parts of a video and proposed an effectiveness curve.

Figure 2 represents the effectiveness of students with a video (axis y) depending on the percentage of video completed by this student (axis x). Line b was proposed by the authors taking into account their own analysis of the videos, while line d is a curve calculated as the mean of the curves proposed by the 8 experts. Both curves are quite similar, as the experts' analysis was equivalent to the one made by the authors, thus validating the proposal of line b. It is noteworthy that this is not a linear function due to the aforementioned characteristics of videos and the hypotheses validated by the experts.

The numpy package[1] was used in order to find a polynomial that fits the curve. The function used to find the polynomial needs to receive as input several coordinates (x, y), and then the function calculates a curve that passes for all these coordinates. In this way, the correspondent effectiveness of the video can be calculated for any percentage of videos the student completes. Next, the non-linear equation for video effectiveness is provided:

Video_effectiveness = $-2.7756 \cdot 10^{-16} + 6.6510x \cdot 10^{-3} - 7.2637x^2 \cdot 10^{-5} + 1.6195x^3 \cdot 10^{-6} - 1.1503x^4 \cdot 10^{-8} + 5.9213x^5 \cdot 10^{-11}$

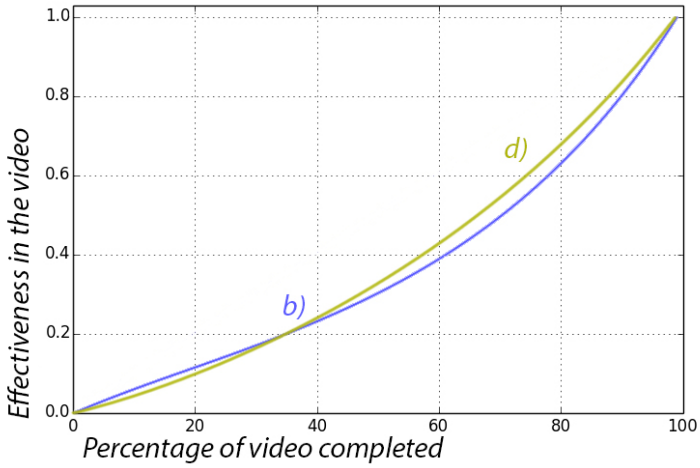


**Figure 2. Polynomial representation to calculate the effectiveness of students with a video in the case study. Line d shows the means of the curves proposed by the 8 experts; Line b is the line corresponding to the authors' analysis of the videos.**

Figure 3 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 was proposed by the authors taking into account their own analysis, while line d is a curve calculated as the mean of the curves proposed by the 8 experts. Both curves are very close, thus validating the proposal of line b. 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. In a similar way as with videos, the numpy package was used to find a polynomial that passes for all the eight different values of number of correct exercises. Next, the non-linear equation for exercise effectiveness is provided:

Exercise_effectiveness = $0.0037 + 0.3196x - 0.0362x^2 + 0.0015x^3$

[1] http://docs.scipy.org/doc/numpy/reference/routines.polynomials.polynomial.html

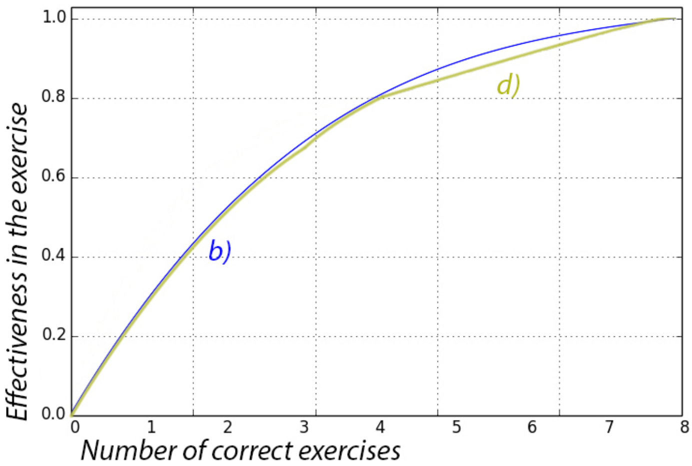


**Figure 3. Polynomial representation to calculate the effectiveness of students with a parametric exercise in the case study. Line d shows the means of the curves proposed by the 8 experts; Line b is the line corresponding to the authors' analysis of the exercises.**

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, the curves proposed by the authors were validated. We then took these curves as the base for the calculation of effectiveness in this research, as these were a result of a more careful analysis of videos and exercises for the four 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.

#### *4.2.4. ANALYSIS OF OTHER TYPES OF ACTIVITIES*

As part of Phase II, we have followed PES for the calculation of the effectiveness of students with videos and exercises. The use of videos and exercises in MOOCs and SPOCs is typically very intensive. However, other activities, such as threaded discussions in social tools (e.g., forums) are pretty common, particularly in MOOCs, where the social component is critical to help peers solve questions and problems. The effectiveness of students with threaded discussions can also be calculated following PES as in the following example.

John is a math teacher that wants his students to reflect about "Rolle's Theorem". He configures a threaded discussion in which each student should explain "Rolle's Theorem" with his own words. Students need to provide an answer before being able to see their classmates' answers and comment on them. In order to automate the quality of the answer, an ontology representing the main concepts in "Rolle's Theorem" is used. The student will get an initial measure of the effectiveness in the interval [0, 1] depending on the concepts included in his answer (according to the predefined ontology). Later on, the student will be able to comment on others' answers. Depending on his comments in the new threads, the system will evaluate the concepts included in his answer (e.g., if the student covers new ones that were not in his first answer, the system will increase his efficiency based on the defined ontology). This way, the system will determine a value for the effectiveness of that student with the threaded discussion.

### 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 and was of the same importance. 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, we 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 metrics can be more or less precise depending on the characteristics of the educational context and the approximations made; the lower the number of approximations, the more the precision obtained in the calculation of the effectiveness. 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 assumption of applying the same function to all the videos and the same function to all the exercises in Phase II.

In addition, even for quite similar videos and exercises, 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 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 and science, that can be more theoretical. In these areas, the effectiveness follows a more linear function".

In the case study, for the case of exercises, the factor that affects the most the resulting metrics based on PES are the relationships among the contents covered in different exercises. The designer of the metrics based on PES should think about what is the increase of knowledge that an exercise might potentially produce on a student when it is solved correctly, taken into account that the student had previously solved correctly some other exercises. A result of effectiveness for an exercise should be provided by the designer for each possible combination of other exercises correctly solved previously. Therefore, for instance, if there are four exercises in a course, designers should provide seven calculations of effectiveness for each exercise (combinations of three exercises taking one by one plus combinations of three elements taken two by two plus all three exercises together). For example, if there are three independent exercises but there is another fourth exercise that covers all the knowledge of the other three exercises, then the obtained effectiveness for just the former exercise should be greater if the three exercises were not solved correctly previously. In our case study, the proposed function for exercises gives this effectiveness implicitly, considering that only exercises that change their parameters are the ones that are related.

In the case study, for the case of videos, the factors that affect most the resulting metrics based on PES are the relationships among different videos, the parts that are more important in a video and the relationships among the different parts within a video. For the calculation of effectiveness, the designer should not take into account the fragments of videos that are not important for learning, and should give less weight to the parts that are not so important. Moreover, the designer should think about what a student can potentially learn for each possible fragment of video, taking into account that the student had watched any combinations of fragments of videos before. In general, this calculation is much more difficult for videos than for exercises because there are infinite possibilities as time watching a video is a continuous variable. For this reason, a detailed analysis of the videos and their relationships is required to detect ways of generalization as in the case study (e.g., there were no unimportant fragments, all fragments were equally important or fragments in a video were all related).

# 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, we carry out a comparison of the global effectiveness between the four SPOCs; 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 behavioral patterns that aim to help teachers make decisions about whether to promote or to avoid certain behavioral 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 4 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 different clusters at a glance:

- 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.

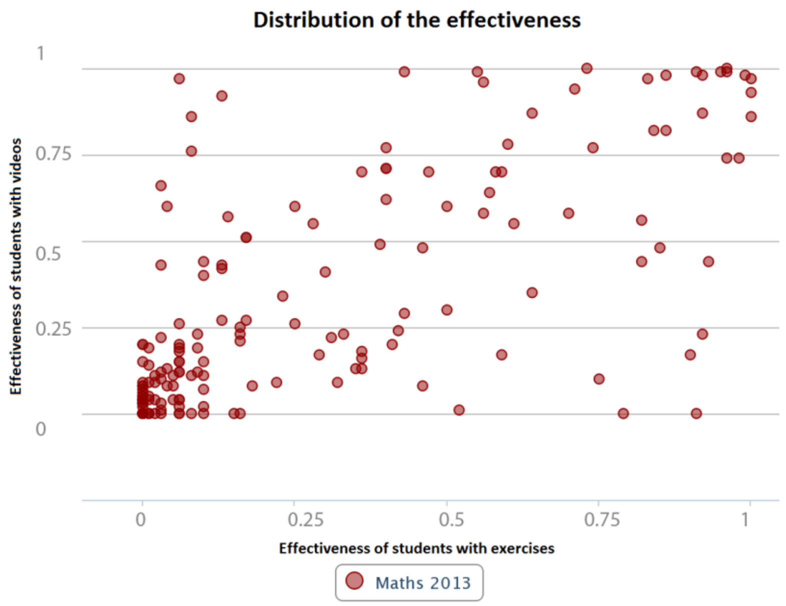


**Figure 4. Analysis of the effectiveness of students with videos and exercises in the SPOC in mathematics.**

Figure 5 classifies the students in Figure 4 in 5 clusters regarding their level of effectiveness with videos and exercises. This visualization can be particularly useful to detect resources and activities that are poorly balanced (high proportion of students in areas tagged "2" and "3"). The clustering of students combines 5 groups:

- Do nothing (neither exercises nor videos) or very little (area tagged "1")
- Do everything or almost everything (area tagged "4")
- Do only or mainly videos (area tagged "3")
- Do only or mainly exercises (area tagged "2")
- Do some videos and some exercises (area tagged "5")

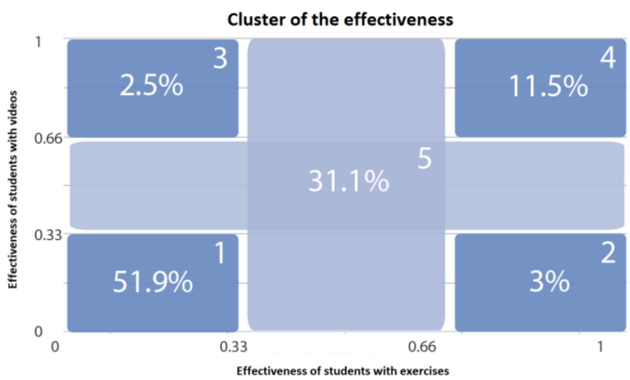


**Figure 5. Clusters representing the effectiveness of students with videos and exercises in the SPOC in mathematics.**

In addition, Figure 6 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 the dropout rate was higher, or in which course students achieved a higher effectiveness with videos or with exercises. It is convenient to note that Figures 4-6 enable the visualization of the effectiveness for two (and eventually three in a 3D representation) kinds of contents at the same time. If the effectiveness of students with other contents is calculated (e.g., animations, social tools, peer reviews, etc.), then the types of contents will need to be combined in dyads or triads in order to use these representations, and further research should be done to clearly plot this information visually.

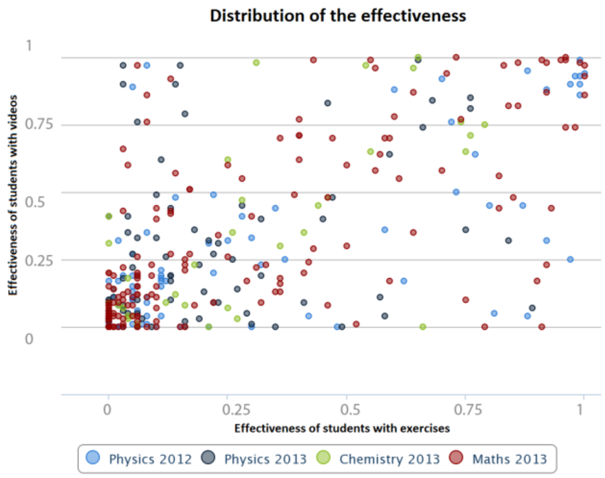


**Figure 6. Analysis of the effectiveness of students with videos and exercises in the four SPOCs.**

Finally, explanations with the criteria followed for calculating the effectiveness of students with resources and activities should be attached to the visualizations. This way, teachers can easily interpret the visualizations and carry out informed decisions.

## 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 7, which compares the four SPOCs under study. It is noteworthy that each point in Figure 7 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 students' effectiveness with videos and exercises. Higher effectiveness in a course does not always mean that this course was designed better, and the specific context should be taken into account. For example, it might be possible that students had a better background in mathematics than in physics from the high school, and so, they do not need to interact so much with videos or exercises. In other occasions, low effectiveness in a course might denote that teachers should 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.

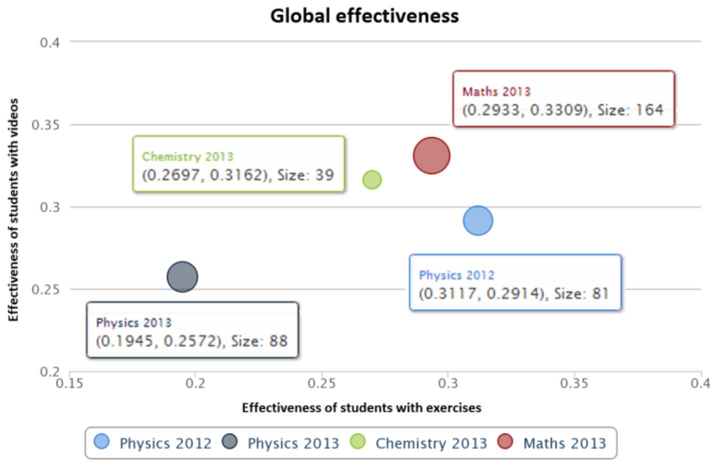


**Figure 7. Comparison of the effectiveness of students with videos and exercises in the four SPOCs.**

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.

Effectiveness of students with exercises was not significantly affected by the type of course, either as can be concluded from the application of the Kruskal-Wallis test, $H(3) = 4.90$, $p > .05$. We use the Kruskal-Wallis non-parametric test in this case as we cannot assume homogeneity of variances and the number of cases of each group is different. As the difference is not statistically significant regarding effectiveness with videos and effectiveness with exercises, then there is no follow-up analysis with post-hoc tests to compare the groups two by two.

| | # 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' behavioral patterns when watching videos and solving exercises (Muñoz-Merino, Ruipérez-Valiente, & Delgado-Kloos, 2013). These metrics are also in the range [0-1] and are:

- *Recommendation listener*. Platforms like the Khan Academy present students the next recommended exercises they can do. These recommendations are based on the prerequisites of each exercise. This metric gives an idea if a student follows the system recommendations about the next exercises to solve.

- *Hint avoider*. Platforms like the Khan Academy offer students the possibility of asking for hints when solving an exercise. Hints are intended to help students, but sometimes students avoid hints and do not select them, even if they are stack solving an exercise. This metric shows if a student avoids selecting hints when he should ask for them in order to solve an exercise.

- *Hint abuser*. On the opposite side, a student can automatically ask for hints without reflecting, that is, even if the student is able to solve the exercise without help. This metric indicates if a student asks for hints when he does not have to in order to solve an exercise.

- *Video avoider*. This metric gives information about if a student avoids watching videos when he should watch them. For example, a student might avoid watching a video related to an exercise, even if he does not know how to solve the related exercise.

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

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<br>(Sig. 0.000) |
| **Effectiveness of students with exercises** | 0.640*<br>(p= 0.000) | 1 |
| **Recommendation listener** | -0.130*<br>(p= 0.012) | -0.137*<br>(p= 0.008) |
| **Hint avoider** | 0.107*<br>(p= 0.039) | 0.032<br>(p= 0.544) |
| **Hint abuser** | -0.092<br>(sig. 0.077) | -0.134*<br>(p= 0.010) |
| **Video avoider** | -0.228*<br>(p= 0.000) | - 0.107*<br>(p= 0.039) |
| **Unreflective student** | 0.048<br>(p= 0.353) | 0.003<br>(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. A reasonable hypothesis is that part of the cause of effectiveness with exercises is effectiveness with videos. In addition, another reasonable hypothesis is that there is a third variable responsible for the effectiveness with exercises: students with a high effectiveness with videos interact a lot in the platform. So, it is more likely that if a student interacts frequently with the platform she/he could have better effectiveness with exercises because at least they are going to try to do the exercises.

There are also several statistically significant relationships between the effectiveness of students with videos and exercises and other students' behavioral metrics at 95% level. In all the cases, correlations are moderate and, in most cases, the correlation is negative:

- **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. Since no prerequisites were set up in the platform among exercises, then the recommendations were random. Therefore, for this case study, no better instruction is given by following these recommendations. One hypothesis for this result is that students who take their own decisions instead of following recommendations can do slightly better in terms of effectiveness

- **Video avoider and effectiveness of students with videos and with exercises.** This relationship indicates that students who avoid watching videos when they are not able to solve the correspondent exercise, are less effective when interacting with the videos they access (as one could expect because they avoid watching videos even when they need to do it, so their video effectiveness will be reduced as they do not watch the correspondent videos in these situations). In addition, they are less effective when interacting with exercises. A reasonable hypothesis is that students have more difficulties to solve correctly exercises that are related with those videos that students avoid. Therefore, this way, their effectiveness decreases.

- **Hint abuse and effectiveness of students with exercises.** This relationship might indicate that abusing of hints has a slightly 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 because students have more information available. On the contrary, a possible hypothesis might be that students who abuse of hints do not reflect on them so they are less able to do the exercises correctly.

- **Hint avoider and effectiveness of students with videos.** This relationship might imply that students who are more effective when interacting with videos will need slightly less hints because they already know the contents.

# 6. Conclusions and Future Work

The level of achievement of the four research questions presented in the introduction is high. First, we solve the first research question with the description of PES, which establishes some general guidelines for the process of defining effectiveness metrics. Among the guidelines provided by PES are to encourage to take into account the different relationships among learning activities, to encourage to take into account the nature of each learning activity, the definition of a continuous range of values for defining effectiveness, or the division of the process in four phases that need to be followed in a given order (selection, effectiveness on each LA independently, effectiveness on a set of LAs of the same type, effectiveness of all LAs). In any case, PES does not define the specific metrics for each learning context, but just the general strategy, as pointed out in the first research question. 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.

Regarding the second research question, PES has been successfully applied to four SPOCs based on video lectures and automatic correction exercises, and deployed in the Khan Academy platform. The process of applying PES for this specific case required following the four phases, and formulas for the calculation were provided. For phase number 2, a careful description of each independent resource was required. Some hypotheses about these resources were formulated and finally a validation by experts of these hypotheses was carried out. In our specific case, metrics based on non-linear functions were proposed for calculating the effectiveness of students with videos and exercises in these courses.

Regarding the third research question, 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. All proposed visualizations aimed at using one axis for the effectiveness on a specific learning resource. Some of the visualizations represented students in a course as points, which enables the comparison of students. When different points from different courses are added, then this allows a fine comparison among students in different courses. There is also the possibility of clustering the 2-dimension space into a visualization and classifying students in categories. In addition, another visualization to compare courses was presented in which just the means of effectiveness of each course are represented with the number of students in each course. This enables the comparison of the global proportion between effectiveness in the different LAs to detect e.g., LAs with lower effectiveness than others. In addition, this visualization allows the comparison of courses as a whole. Statistical significance of the differences can be calculated with the proper statistical tests. There is a need of analysis of the specific learning context to reach conclusions about the differences in effectiveness.

Regarding research question 4, 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 run. In addition, some interesting relationships among the effectiveness metrics defined in these case study and other behavioral metrics published in the literature were found. Reasonable hypotheses about the causes of some of these relationships have been formulated.

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 Khan 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. It is important to note that effectiveness on videos and exercises is not equal to effective learning. New experiments including pre and post-tests to analyze students' learning gains will help on re-shaping these metrics and finding relationships between effectiveness of students with resources and activities and learning 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. As part of the future work, we also aim to evaluate the usefulness of these visualizations with teachers and other stakeholders (e.g., institution decision makers). Also, we will explore and study how to implement visualizations in dyads or triads when adding more than two or three variables to the graph.

In conclusion, 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.

## ACKNOWLEDGEMENTS

This work has been supported by the EEE project (Spanish Ministry of Science and Innovation, "Plan Nacional de I+D+I TIN2011-28308-C03-01), the "GEEWHEZ: ManaGEmEnt and leisure middleWare for tHemE parks and Zoo" project (European Commission under grant FP7-SME-2011-286533) and the postdoctoral fellowship Alliance 4 Universities.

We would like to thank the professors in Universidad Carlos III de Madrid who developed the Physics, Mathematics and Chemistry materials for the courses (videos and exercises).

## REFERENCES

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

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

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

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

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

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

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

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

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

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

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

Hiltz, S.R., & Arbaugh, J.B. (2003). 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, 59-72.

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

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

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

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

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

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

Pardos, Z.A., Baker, R., San Pedro, M., Gowda, S.M., & Gowda, S.M. (2013). Affective states and state tests: investigating how affect throughout the school year predicts end of year learning outcomes. In *Proc. LAK 2013*, 117-124.

Price, G. (2013). *Are We Beginning the Post-MOOC ERA? Two SPOCs (Small Private Online Courses) are Underway From Harvard/edX*. <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>.

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

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

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

Suebnukarn, S., & Haddawy, P. (2006). A Bayesian approach to generating tutorial hints in a collaborative medical problem-based learning system. *Artificial Intelligence in Medicine*, 38, 5-24.

Swan, K. (2003). 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, 13-45.

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

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

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