A Study on the Pedagogical Components of Massive Online Courses

Un estudio sobre los componentes pedagógicos de los cursos online masivos

Dr. Manuela Raposo-Rivas is Professor in the Didactics in the School Organisation and Research Methods Department of the Faculty of Education at the University of Vigo, in Ourense (Spain) (mirapo@uvigo.es).

Dr. Esther Martinez-Figueira is Professor in the Didactics in the School Organisation and Research Methods Department of the Faculty of Education at the University of Vigo, in Ourense (Spain) (esthermf@uvigo.es).

Dr. José-Antonio Sarmiento Campos is Associate Professor in the Didactics, School Organisation and Research Methods Department of the Faculty of Education at the University of Vigo, in Ourense (Spain) (sarmiento@uvigo.es).

ABSTRACT
At the end of the first decade of the 21st century, the rapid growth in the MOOCs offering brought about a new educational landscape, posing new challenges to teaching and learning, mainly due to massive participation, ubiquity and free enrollment. These courses embody a confluence of technological and pedagogical mediations yet to be fully explored. This study, applying an instrument of educational and interactive indicators on a MOOC (INdiMOOC-EdI) designed by the authors, seeks to ascertain the pedagogical components of MOOCs delivered in the Spanish language in order to establish which course features are platform dependent. An exploratory sequential mixed methods study was carried out on a total of 117 courses within 10 different platforms via an instrument developed and validated ad hoc. The data obtained is subjected to a content analysis for a qualitative perspective; while a quantitative perspective is obtained through a statistical analysis complemented with classification algorithms typical in data mining. The results suggest that the different platform providers condition the pedagogical design of the MOOCs in five key areas: learning, activities and tasks, means and resources, interactivity and assessment. It concludes with a series of descriptive, educational and interactive indicators that can be a pedagogical benchmark for future MOOCs.

RESUMEN
A finales de la primera década del siglo XXI, el rápido aumento de cursos MOOC dibuja un nuevo panorama educativo planteando nuevos retos a la enseñanza y al aprendizaje debido, fundamentalmente, a sus características de masividad, ubicuidad y gratuidad. En estos cursos se da una confluencia de mediaciones tecnológicas y pedagógicas aún por explorar en todas sus dimensiones. Este trabajo, utilizando un instrumento de indicadores educativos e interactivos en un MOOC (INdiMOOC-EdI) de elaboración propia, se centra en averiguar qué componentes pedagógicos posee la actual oferta de MOOC en lengua hispana, para poder dirigir aquellos elementos dependientes de las plataformas que las soportan. Para ello se realiza una investigación mixta de tipo exploratoria y secuencial que analiza un total de 117 cursos ubicados en 10 plataformas diferentes, utilizando el instrumento creado y validado ad hoc. Con la información obtenida se lleva a cabo un análisis de contenido en su vertiente cualitativa, mientras que con la cuantitativa se efectúan análisis estadísticos complementados con algoritmos propios de la minería de datos. Los resultados muestran que las diferentes plataformas condicionan los diseños pedagógicos del MOOC en cinco aspectos fundamentales: el aprendizaje, las actividades y tareas, los medios y recursos, la interactividad y la evaluación. Se concluye con una serie de indicadores de tipo descriptivo, formativo e interactivo que pueden orientar la pedagogía de futuros Cursos Online Masivos Abiertos.

KEYWORDS | PALABRAS CLAVE
Online courses, Massive Online Open Courses, MOOC, open educational resources, instructional design, data mining, content analysis, web-based instruction.

Cursos en línea, Cursos Online Masivos Abiertos, COMA, recursos educativos abiertos, diseño instructivo, minería de datos, análisis de contenido, formación basada en la web.
1. Introduction

MOOCs are a phenomenon of utmost interest to the scientific community due to their exponential growth (Liyanagunawardena, Adams & Williams, 2013; Martínez, Rodríguez & García, 2014; Yuan & Powell, 2013). These courses are a worldwide expanding phenomenon and offer a clear example of disruption (Anderson & McGreal, 2012; Conole, 2013; Vázquez Cano-Lopez & Saracola, 2013) due to low fees for participants, massive participation and their adaptation to new social needs regarding education. Hence, the disruptive nature of MOOCs can only be verified if taken as experiments to test new methodology, technology and new ways to organize education (Pernías & Luján-Mora, 2013).

From a pedagogical point of view, the phenomenon can be seen as an 'effervescence' rather than a disruption (Roig, Mengual-Andrés & Suarez, 2014), which must not blind us to the reactions they stir. These courses are hosted by varied and diverse platforms, with different backgrounds and approaches which have given rise to MOOCs based on: web-based instruction, the Connectivist Theory and its pedagogical model (Siemens, 2005); tasks, according to competency-based accomplishments (Cormier & Siemens, 2010) and content (Pernías & Luján-Mora, 2013; Vázquez-Cano, 2013).

Since the inception of MOOCs, the majority of studies have focused mainly on their concept and history of MOOCs; debating the challenges, possibilities and threats thereof; presenting case studies by examining one or more platforms and courses, and reflecting on student participation (Liyanagunawardena et al., 2013). Thereafter, the focus shifted mainly to the completion rates and course quality per se (Baxter & Haycock, 2014; Halawa, Greene & Mitchell, 2014; Jordan, 2014; Koutropoulos et al., 2012; Rodríguez, 2012) and their pedagogical principles (Glance, Forsey & Riley, 2013; Roig et al., 2014; Vázquez-Cano et al., 2013; Zapata, 2013); design and key components are scarcely addressed.

Despite the lack of consensus on how quality standards should be attained in MOOCs (Haggard, 2013), it is necessary to raise the issue in order to prevent MOOCs from becoming poor quality video watching sessions of chatting professors which are the basis for a set of self-assessment questions and awarding certificates without prior authentication and no other concern except generating revenues (Aguaded, 2013: 7-8).

It is therefore important to address what pre-course information is provided, the pedagogical approaches underlying the design, the level of student engagement, the role of course instructors, availability and degree of interaction, resource typology as well as certification structure and process (Vázquez-Cano, 2013; Zapata, 2013).

Research on these training approaches shows that they are founded on a decentralized control over teaching-learning processes (Baggaley, 2014). However, given the accessibility and reach of MOOCs there is almost by definition a wide spectrum of users with a variety of interests and motivations, approaches and learning styles; hence, one of the most difficult challenges is to provide authentic learning experiences, which require the design and development of interactive collaborative processes. Siemens (2005) states that cooperative and collaborative activities as well as interaction with technological resources have a direct impact on students, especially on the way they perceive and process information and on their learning process, thus prompting a new knowledge building approach. Given massive student participation the level of interactivity is addressed through the use of specific Web 2.0 collaborative and communicative tools: chat rooms and forums (Baxter & Haycock, 2014) to discuss concerns and share solutions; blogs, wiki-forums and social networks (Medina-Salguero & Aguaded, 2014), among others, for support and feedback.

Assessment normally conforms to formal and summative processes that are determined by the type of accreditation awarded once the MOOCs has been successfully completed. In some cases, the objectives are small-scale goals carried out individually or in pairs which are assessed by means of surveys, questionnaires, quizzes, exams, problem sets and other processes that will automatically generate badges as evidence of learning.

In short, studies have focused on the characteristics of the platform providers and the success or failure of a given course (Fini, 2009) and less on the pedagogical aspects. If we want to maximize learning via analyzing and adapting teaching strategies to individuals, we must critically address the pedagogical design of the MOOCs to identify underlying trends in teaching and learning processes. On the basis of the aforementioned, the objectives of this research are:

• To analyze Spanish-language MOOC offering during a given period in order to establish a profile of the pedagogical components.
• To validate a tool that can guide the pedagogical design of MOOCs.
• To ascertain which components are unique to a MOOC from those dependent on the platform.
2. Material and methods

The purpose of this research is descriptive with an exploratory sequential mixed-methods design (DEXPLOS) (Creswell, Plano, Gutmann & Hanson, 2008; Hernández, Fernández & Baptista, 2010). This design involves an initial phase of qualitative data gathering and analysis followed by another where quantitative data is collected and analyzed. Subsequently, we create another database that integrates both and enables mixture analysis techniques (García, 2011).

Sequencial and criteria sampling (McMillan & Schumacher, 2005) for mixed methods (Hernández & al., 2010) is used. The courses were selected according to the following criteria: catalogued in the repository www.MOOC.es; delivered in Spanish; course information available without prior registration; and provide a minimum amount of information to the data collecting instrument.

We therefore focus on ten platforms (Open UGR, Coursera, MiríadaX, Tutellus, Ucam, Udemy, UnedComa, Unimooc, UNX, UPVX. We discarded Ehusfera (a blog hosting service rather than a MOOC platform) Iversity, CourseSites and edX, among others, given that the reference language is not Spanish. This involved analyzing 117 courses from different fields of knowledge available during the month of March 2014 (table 1).

The low percentage of courses from Tutellus and Udemy is mainly due to two factors:

- They included material that did not conform to the MOOC concept, such as conferences, videoconferences or lectures on videos, recycled from different sources within the audiovisual repository of the institution and now offered as massive courses.
- They provided very limited information to the research instrument without prior registration. Moreover, there was redundant information on how to use the platform and on certification. It was also noted that there was a high degree of repetition, such that regardless of the course, the data provided was the same.

Consequently, these two platforms were not included in the qualitative sample. Thus, we can state that the remaining 104 courses represent 81.25% of the population.

2.1. Instrument

For massive course analysis we developed INdiMOOC-EdI (Instrument for Educational and Interactive Indicators in MOOCs). It is an ad hoc data sheet that meticulously collects information provided in the full description of MOOCs. The elements that make up this instrument can be organized into four components, with a total of 27 sub-components rated on

<table>
<thead>
<tr>
<th>Table 1. Participating Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Platform</strong></td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>Albert@UGR</td>
</tr>
<tr>
<td>Coursera</td>
</tr>
<tr>
<td>MiríadaX</td>
</tr>
<tr>
<td>Tutellus*</td>
</tr>
<tr>
<td>Ucam</td>
</tr>
<tr>
<td>Udemy*</td>
</tr>
<tr>
<td>UNED-COMA</td>
</tr>
<tr>
<td>Unimooc</td>
</tr>
<tr>
<td>UnX</td>
</tr>
<tr>
<td>UPVX</td>
</tr>
</tbody>
</table>

*Courses discarded in the qualitative sample.
To safeguard validity conditions, the first version of the instrument was subjected to the Delphi technique by evaluating the same courses during the same period of time and a pilot study of 15 courses within 5 different platforms. The expert panel (KC) index rated .75, while content validity (IVC) rated .99 which according to Lawshe (1975) is within the standard satisfactory level. Reliability and internal consistency were determined by Cronbach’s Alpha statistic after the sample gathering procedure i.e., once questions whose answers were measured on an interval scale were eliminated. The 117 courses obtained an alpha value of 614. Some authors (Huh, Delorme & Reid, 2006; Nunnally, 1967) indicate that an alpha value between .5 and .6 is within the satisfactory standard in the early stages of research or in an exploratory study such as this one. This statistic combines the correlation coefficient of the items that make up the instrument and its dimensionality (Cortina, 1993).

3. Analysis and results

A descriptive analysis of quantitative data was carried out according to the identifiers and descriptive features displayed in table 2, together with a categorical principal component analysis that enables a large set of variables to be grouped in a smaller number of explanatory components that stem from the variance among the original data.

With the qualitative data (interactive and educational features) we conducted a content analysis that deployed five major categories:

- Learning: styles, learning modality taking place and content format: self-directed learning, empirical and inductive learning, learning through observation; lessons, units, pills or modules.
- Activities and tasks: refers to both modality (compulsory or optional, individual or collective) and typology (questionnaires, tests, readings, practical exercises, problem sets, projects, case studies, questions and answers, participation…).
- Means and resources: traditional and technology-based: videos, slideshows, forums, blogs, wikis, e-mails, interviews, readings, optional additional material.
- Interactivity: or interacting with other people; online meetings, debates or discussions in pairs or groups, sharing doubts and knowledge, collaborative work, flexible and asynchronous communication.
- Assessment: existing assessment procedure, not only modalities and instruments, but also grades and endorsement (self-assessment, peer assessment; questionnaires, tests, rubrics, exams, problem solving, –peer-peer); percentage or weighting in the final grade, grading scale, passing grade, minimum percentage; checking student progress and final endorsement.

Figure 1 shows the relationship between categories and associated codes, taking into account that some codes belong to more than one category.

Two algorithms of data mining were applied, which will later be described in detail herein: first, a classification algorithm to discern the impact of platforms on the instructional and communication designs underlying the courses; second, an assessment algorithm to ascertain the degree of information provided by the variable course regardless of the platform provider.

3.1. What profile do Spanish-language MOOCs present?

The analysis conducted with the information compiled reveals that the 98.3% of the courses (n=115) have the title somewhere clearly visible, crucial in order to engage participants’ interest, as well as a limited registration period (n=38, 32.5%) or unlimited registration period (n=34, 29.1%). In 38.5% of the cases (n=45) registration was closed during the study timeframe.
A total of 72.6% (n=85) are sponsored by platforms linked to Higher Education; whilst personal initiative (n=13, 11.1%) or private company sponsors (n=1, 0.9%) are less frequent. As far as fields of knowledge are concerned, almost half of the MOOCs relate to Legal and Social fields (n=49, 41.9%), followed by multidisciplinary MOOCs (n=21, 17.9%), Arts and Humanities together with Science MOOCs (n=15 each, 12.8%). The least offered are Technology (n=10, 8.5%) and Health Sciences MOOCs (n=7, 6%).

Courses analyzed, only n=49 (41.9%) of the participants specify course relevance. More than half (n=63, 53.8%) lack addressee information. When addressing target participants, n=34 (29.1%) they note for public at large and n=20 (17.1%) establish a specific profile. In almost 60% of the courses (n=70, 59.8%) there are no prerequisites. Regarding the two last issues there are five important aspects that prompt registration:

- Including an Introduction to the course in the MOOC website. Almost half of the introductions deal with content (n=47, 40.2%), followed by 38.5% (n=45) which focus on the topic, without being too concise. The rest (n=25, 21.4%) address issues such as timing, objectives, using the system, carrying out tasks, etc.
- Having an introductory video, available in practically all of the courses analyzed (n=98, 83.8%).
- Having and defining objectives is omitted in more than half of the courses (n=67, 57.3%).
- MOOC related courses were only available in a small percentage of the courses (n=38, 32.5%)
- Operation of the system is specifically addressed in the platform in most MOOCs (n=91, 77.8%), only in 9.4% (n=11) is this guidance provided via course. In n=14 (12%) it is not specified.

The length of the MOOC analyzed is normally limited to weeks (n=87, 74.2%) ranging from 6 weeks (n=22, 36.7%) to 7-8 weeks (n=19, 31.7%). Thus, unlimited course length is a rare occurrence (n=11, 9.4%). Furthermore, the duration of engagement is specified in n=83 (79.9%), generally ranging from 3 hours (n=28, 46.7%) to more than 5 hours per week (n=19, 31.7%). Less than 2 hours of weekly engagement is infrequent (n=5, 8.3%).

A high percentage (n=84, 71.8%) of courses present the MOOC teaching team in a visible area, with an average of 3 to 4 tutors (M=3.32 and SD=3.148). This information is not displayed in only 17.1% (n=20) and the remaining courses (n=11, 9.4%) provide no information at all.

Regarding course content there is a tendency to adopt an open structure, lessons or modules (n=90, 76.9%) with an average of 8 modules per MOOC. Less frequently (n=22, 18.8%), it appears that the work plan is limited to weeks only in closed structure courses. There is no information available in n=5 courses (4.3%).

As for certification, there is a combination (n=75, 64.1%) of free of charge and charge-bearing modalities. Regarding the type of accreditation, it is normally mixed (n=71, 60.7%), certificates, credentials, badges, medals, and so on.

The content analysis resulting from the five categories (activities and tasks, learning, assessment, interactivity, means and resources) previously mentioned displays the given trend within each platform (figure 2).

Regardless of the number of courses within each platform, it is noted that Coursera offers higher quality information with regard to educational and interactive features, followed by MiriadaX and UNED-COMA. On the other hand, and except in the aforementioned three platforms, it is observed that platforms are more vulnerable to and deficient in features such as means and resources, activities, tasks and assessment.

3.2. Do platforms condition pedagogical designs?

To address this issue a categorical principal component analysis (CATPCA) is carried out, which is non-linear and therefore does not require the strict assumptions of principal component analyses (Molina & Espinosa, 2010), regarding two dimensions as necessary and sufficient to yield representation (figure 3). The data obtained confirms that the amount of variance accounted for by these two dimensions is not high (s²=10.64%), but underscores a substantial difference among courses within different platforms. In the first dimension the saturating variables are: certification (.943), engagement (.905), dedication hours (.899), accreditation (.864), registration (.872) and institution (-.883). The variables that saturate the second dimension are: introductory video (-.717), teaching team (80,625), faculty profiles (.608) and modules (.629). Although there are variables that do not cluster significantly in any dimension, it is true that the vast majority have opposing values in one or another.

Plotting the two dimensions in a coordinate axis displays how courses are grouped according to the platform provider. The outcome reveals the following facts:

- Some platforms are more extensive than others, for instance UniMOOC is within the values x=-2 and
y=2, whilst MiriadaX extends from x=-1.
• There is certain affinity among platforms, which can lead to conglomerates, inter alia, Coursera and MiriadaX, UniMOCC and UPVX...
• In the sample studied the pedagogical components of each course are fully determined by the platform that hosts them.

To give an in-depth insight, the data collected were subjected to an algorithm classification with Weka software (Hall & al., 2009). Since we are dealing with a collection of automatic learning algorithms for data mining tasks, the platforms under investigation are regarded as the classification variable and producing ten rules that classify 100% of the courses in a platform. The algorithm used was PART (Frank & Witten, 1998), a variant of the C4.5 of Quinlan (1993). As an example we provide a fragment of three rules:

The outcome demonstrates that the relevance of the platform is greater than that of each course when it comes to the pedagogical design. For instance, we observed that in MiriadaX, where more courses were analyzed, both types of certification, dedication and limited course length, together with not displaying related courses, are associated with this platform (n=55.0, i.e., all the courses examined).

3.3. Which MOOC components are platform-independent?

To examine what components are more specific to the course than to the platform, the data were subjected to an algorithm implemented in Weka which assesses the rate of each attribute by measuring the information gain (Witten, Frank & Hall, 2011) according to java class platform (table 3).

It is noted that most of the information provided by the variables is related to the platform. The title and interactivity variables do not display any variance; hence it can be attributed neither to course or platform. Figure 4 shows the values obtained through the algorithm once standardized. If negative values are to be taken as typical course variables, the following should be considered: Field, Introductory Video, Introduction, Target participants, Modules, Objectives, Teaching team, Importance to the public, Prerequisites and Lenght/weeks. If we increase to a DT = -1, the typical course variables would be Weeks and Prerequisites.

4. Discussion and conclusions

An overview of literature addressing MOOCs emphasizes the relevance of pedagogy in MOOCs. The use of the instrument (INdiMOOC-EdI), which enables an analysis of these components, was contemplated and implemented and has been applied to a total of 117 Spanish-language courses.

The study on the data obtained with the aforementioned instrument regarding descriptive, educational and interactive features determines that it can be employed as a benchmark of indicators in order to
attain the desirable pedagogical design in a MOOC. Some of the findings affirm, along with Glance & al. (2013), that MOOCs have a sound pedagogical basis and there is no reason to assume that they are less effective than other learning experiences.

Initially, the variable course is analyzed; however, the data refer to the support platforms. The information obtained underscores a series of components that are not persistent in the vast majority of the courses analyzed. For instance, platforms do not regard it essential to specify: target participants, prerequisites and a clear and concise introduction about the course content or other related courses. Nonetheless, it does seem necessary to display: an introductory video; objectives, teaching team, length and weekly dedication, operation of the system, means and resources, activities and assessment. It has been shown that assessment is one of the most evident pedagogical benefits of MOOCs (Glance & al., 2013).

The existing profile of Spanish-language MOOCs, drawn from the pedagogical features in more than 70% of the courses examined, includes key features such as displaying the course title in a visible place (98.3%); an introductory video (83.8%); specifically addressing operation of the system (77.8%); an open structure, modules or lessons (76.9%) with an average of 8 modules per MOOC; course length limited to weeks (74.4%); platform provider linked to Higher Education (72.6%), displaying the teaching team in a visible spot (71.8%) and specifying the number of hours of weekly dedication (70.9%).

Although the platform offers the necessary technological support, it is obvious that MOOC proposals should have their own autonomy. Contrary to what is asserted in the study by Roig & al. (2014), which found no significant variance between the pedagogical quality of MOOC and the platform, it is ascertained that platforms determine the pedagogical design of the courses. The same pattern comprising, inter alia, activities and materials, learning modalities, assessment proposals, level of interactivity, access and certification is repeated over and over again (as many times as the number of platforms examined). The characteristics of the activities students carry out along with tutor counselling and didactic interventions are key elements in predicting the rate of disengaging and drop-outs (Halawa & al., 2014).

The data implies that platforms condition the pedagogical designs of MOOCs (figure 3), but this does not necessarily imply the existence of a pedagogical model underlying the MOOC proposal. I.e., the platform constrains and restricts online courses, albeit some platforms deploy a degree of flexibility, with fluid boundaries among the different features in INdI-MOOC-EdI.

If MOOCs are regarded as a dynamic and global phenomenon, as an educational response to the emergence and development of movements and online social networks, as a cybernetic alternative to learning without frontiers, as a useful self-directed learning experience, as an extension of the classroom, as a space for free movement of knowledge, as an opportunity for democratization and universal access to specialized content, as a training proposal with pedagogic autonomy... then let us take advantage of these mentoring platforms whilst MOOCs have not fully matured.

Finally, an exploratory study such as the one carried out provides an outline of the situation, but it is faced with certain constraints that should be addressed in future research, such as an in-depth insight into the field or methodological complementariness. It would be convenient to thoroughly examine a specific course or courses in specific fields of knowledge; the standpoint in other languages; or if low completion rates

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Information Gain</th>
<th>Attribute</th>
<th>Information Gain</th>
<th>Attribute</th>
<th>Information Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accreditation</td>
<td>0.2289</td>
<td>Importance to the public</td>
<td>0.4223</td>
<td>Introduction</td>
<td>15.9703</td>
</tr>
<tr>
<td>Dedication hours</td>
<td>0.9231</td>
<td>Course length</td>
<td>0.9966</td>
<td>Field of knowledge</td>
<td>0.9966</td>
</tr>
<tr>
<td>Objectives</td>
<td>0.4483</td>
<td>Field of knowledge</td>
<td>0.9966</td>
<td>Faculty Profiles</td>
<td>0.9966</td>
</tr>
<tr>
<td>Certification</td>
<td>0.3637</td>
<td>Prerequisites</td>
<td>0.9966</td>
<td>Recipients</td>
<td>0.9966</td>
</tr>
<tr>
<td>Operation of the system</td>
<td>0.3035</td>
<td>Related courses</td>
<td>0.9966</td>
<td>Title</td>
<td>0.9966</td>
</tr>
<tr>
<td>Teaching team</td>
<td>0.4423</td>
<td>Introductory video</td>
<td>0.9966</td>
<td>Dedication</td>
<td>0.9966</td>
</tr>
<tr>
<td>Registration</td>
<td>0.9966</td>
<td>Interactivity</td>
<td>0.9966</td>
<td>Modules</td>
<td>0.9966</td>
</tr>
</tbody>
</table>

© ISSN: 1134-3478 • e-ISSN: 1988-3293 • Pages 27-35
can be due to the pedagogical design. As stated by Bartolomé (2013), we still lack a pedagogical framework that will validate that a MOOC teaches and that a MOOC generates knowledge. Further research is needed for ongoing progress and consolidation.

There is as yet the need to debug concepts, models and experiences,... overcome certain difficulties and minimize others; some MOOCs and platforms will lag behind, but many others will continue to be designed, developed and improved for millions of people around the world.

Notes
1 The study was carried out during the last academic year in response to an institutional innovation project commissioned to the research group which includes the authors of this paper.
2 Despite the evident educational value objectives possess, they are included in the descriptive features in order to simply determine whether they are present or not, since according to Roig & al. (2014: 37): ‘The existence of explicit learning objectives is associated with a high score in the pedagogical quality of MOOC’s’.
3 Information Gain=Class Entropy - Entropy (class / attribute)= Class H-H (Class /Attribute).

References


