This is a pre-print
Please cite this article as follows:


**METHODODOLOGICAL APPROACHES IN MOOC RESEARCH: RETRACING THE MYTH OF PROTEUS**

**Abstract**

This paper explores the methodological approaches most commonly adopted in the scholarly literature on Massive Open Online Courses (MOOCs), published during the period January 2008 - May 2014. In order to identify trends, gaps and criticalities related to the methodological approaches of this emerging field of research, we analyzed 60 papers selected across four relevant scientific databases plus one journal in the sector of eLearning that published a special issue on this topic. The emerging picture is that of a research field in its infancy, heavily relying on theoretical research and case studies, which is just beginning to identify suitable methods to deal with large cohorts of learners, very large amounts of data and new ways of learning. The state-of-the-art is also quite fragmentary, due to the different epistemological and ontological conceptions of the authors of the papers about the nature of the issues faced and the way they should be studied. In this paper, we compare the problems related to the definition of the methodological approaches in this research field with the Greek myth of Proteus, the elusive, multiform divinity whose wisdom would only be revealed to those capable of catching him. We therefore advocate the need for catching Proteus, that is, better identifying and choosing the methodological approaches for MOOC research as a condition for its development.

**Keywords**

Educational research methodology, Massive Open Online Courses (MOOCs), research methods, eLearning research, Gorard’s cycle of educational research.

1. **Introduction**

Nowadays, there is concern regarding the poor definition of the methodological approaches adopted in much of the educational literature. A methodological approach is the lens through which a phenomenon is observed: unsuitable choices or gaps in terms of the methods used in educational research can lead to unreliable or insufficient research results and conclusions (Gorard & Cook, 2007). Conflicting definitions of topics like teaching and learning effectiveness, the lack of coherence between the definition of research problems and the methods adopted to explore it and the profusion of research methods leading to almost unmanageable masses of data are some of the problems pointed out (Bliuc, Goodyear, Ellis, 2007; Trifonas, 2009; Carr, 2007). The result is a lack of credibility of educational research and
poor transferability of research results to policy making and practice (Bates, 2014; Gorard & Cook, 2007).

The debate about the methodological challenges regards the broader context of educational research, often considered conflictive (Hammersley, 2007). As Carr (2007 p.273-274) affirms: “the sad truth is that educational research now embraces so many traditions, paradigms, theoretical perspectives, methodological frameworks and academic disciplines that it cannot claim to meet even the most minimal criteria of homogeneity that any notion of a “research community” presupposes and requires”.

One of the concerns identified within this debate is the coexistence of “incompatible versions of what research is” (Carr, op.cit, 274), due to incoherent underlying visions of education deriving from “incidental” reflections in diverse fields such as philosophy, psychology and sociology (Stenhouse, 1981, p.113, cited by Carr). By contrast, disciplinary and methodological diversity and complexity are also seen as part of the identity of educational research to “avoid the reductionism of social science [...] via the radical pluralism of bricolage” (Kincheloe & Berry, 2004, p.X). In both the mentioned viewpoints, the authors advocate the need for deeper reflections on methodological issues within the educational research community, in order to overcome internal inconsistencies and improve the scientific status of the field.

The weakness that Carr undelines can also be found in eLearning research, a multi/inter and trans-disciplinary field (Conole & Oliver, 2007), considered as a young branch of educational research (Bates, 2014). Methodological approaches in eLearning research reflect the great diversity that characterizes the heterogeneous nature of problems in the broader field of educational research. Moreover, research in eLearning has been considered rather complex, given the need to take into account the challenges imposed by the hybrid nature of eLearning, thus requiring inventive approaches where conventional ones may not be adequate (Andrews & Haythornwaite, 2007). According to Friesen (2009), this area of educational research is not mature and seems to be strongly linked to specific researchers’ agendas.

Within the wider framework consisting of methodological issues in educational research and, more specifically, in eLearning research, the research field of Massive Open Online Courses (MOOCs) has peculiarities that require the adoption of novel research methods because the problems faced are rather new: large cohorts of students, very large amounts of data, new ways of learning determined by unprecedented freedom of choice. The MOOC phenomenon is often seen as an opportunity to answer compelling questions in the educational agenda, like renewing higher education, so research in this area might both reflect problems similar to those of eLearning, but it could also offer hints to straighten up some of the issues thanks to the narrower, younger and disrupting nature of the field.

In this paper, we explore the methodological approaches most commonly adopted in the scholarly literature about MOOCs published during the period January 2008 - May 2014, in order to understand trends, gaps and criticalities relating to the methodological concern in this field of research and investigate whether the issues affecting research in education and eLearning are inherited by and possibly exacerbated in this specific field.

Characterizing a research study in terms of the methodological approach it adopts is far from being easy: there is no agreement on one single way to classify a piece of work in educational research and most of the classifications proposed tend to be rather abstract and difficult to use in a pragmatic way because the borders between different categories are often blurred.

It is for this reason that our effort in this paper recalls the myth of Proteus.

According to Greek mythology, Proteus, called by Homer in the Odyssey the “Old man of the Sea”, was a divinity that treasured the power to see all things. For the disgrace of those who wanted to take advantage of such a precious ability, Proteus changed shape, becoming unstable and fluid like water. The answers Proteus gave to the lucky captors were sharp and wise, worth the efforts made.
According to Hegel’s interpretation (1817[1991]) of this myth, Proteus represents the Nature and Human soul “in their perpetual transformation”; therefore, catching Proteus means leading the “perpetual transformation” of reality into stable forms. Following Hegel’s words, for us this myth is a metaphor of the attempt to bring order out of the chaos of methodological approaches in MOOC research: Proteus represents such approaches since they are changing and liquid due to the evolving nature of this research field. The researchers try to “capture” the essence of a research problem through a specific methodological approach.

In this article we aim to support educational researchers in their endeavor to “catch Proteus” by: a) depicting a stable panorama of methodological approaches used in MOOC research as a mean of promoting awareness of methodological issues within the research community; b) classifying and sorting the resulting picture and c) promoting a better understanding of methodological approaches thus providing a basis for a sound development of the research field.

2. MOOC research so far

MOOCs have had an impressive coverage in press and blog articles since this “educational experiment”, as Siemens termed it (Siemens, 2012), has started to be seen as a leverage to renew Higher Education (EDUCAUSE, 2012; Knox, Bayne, MacLeod, Ross, Sinclair, 2012). The first scholarly publications appeared in 2009 and focused on the analysis of the MOOC model from an institutional point of view and on its disruptive potential as a new business model in Higher Education. Subsequently, the number of papers significantly increased and interest moved onto the impact that MOOCs have on learners across diverse learning cultures (Siemens, Irvine, & Code, 2013; Mor & Koskinen, 2013; Ullmo & Koskinen, 2014). Two literature reviews have been published on this topic before May 2014 (Ebben & Murphy, 2014; Liyanagunawardena et al. 2013). In the former, two phases of scholarship about MOOCs have been identified (Ebben & Murphy, 2014). The first phase, running approximately from 2009 to 2011, focuses on experimentalists, participants’ enthusiasm and engagement; claims and results were linked with connectivist pedagogical approaches and with Open Learning discourses. The second phase, from 2012 to 2014, focuses on exploring and applying pedagogical and technological approaches aimed at improving access to education and renewing higher education business models. This second phase leaves room for some criticism regarding aspects like drop-out rates (Hill, 2013), cultural translations (Nkuyubwatsi, 2014) and sustainability (Yuan, Powell, & Olivier, 2014). According to the latter review (Liyanagunawardena et al., 2013), four major problems must be faced by MOOC researchers and designers: a) the need to explore all of the stakeholders’ perspectives in MOOCs (learners, creators, teachers, institutions, etc.), taking into consideration advantages and disadvantages for every perspective; b) the cultural tensions within MOOC pedagogies, resources, and learning environments; (c) the ethical concerns about using data generated by MOOCs; and d) the learners’ effective strategies to deal with information overload and self-regulate within MOOCs (including avoiding drop-out).

So far, however, little attention has been paid to the methodological approaches used in research on MOOCs. Some of the challenges regarding this topic have to do with the novelty of the pedagogical problems faced, such as students’ attrition in open and massive online courses, the need for new criteria in course effectiveness analysis and new methods to sustain participants’ motivations and engagement (Coffrin, Corrin, de Barba, & Kennedy, 2014), and methods and tools to collect and analyze large amounts of data concerning learning (Bieke, De Laat, Teplovs, & Voogd, 2014). Nevertheless, a comprehensive picture of the methodological approaches adopted so far would be a sound basis to better characterize educational problems and solve them in a less tentative way. Moreover, the diversity of methodological approaches needs to be
explored in order to promote researchers’ reflection on the gaps and criticalities in the definition of topics, research aims, use of methods for data collection and analysis. In brief, this is precisely what we mean by “catching Proteus”: with the analysis of the methodological approaches adopted in the examined papers, we have tried to lead the “perpetual transformation” of MOOC research into a more stable picture.

3. Methodology

This paper reports on a systematic literature review, based on the analysis of 60 papers of relevant scholarly literature published during the period January 2008 - May 2014. As previously mentioned, the analysis was driven by the objective of drawing a picture of the methodological approaches used so far in MOOC research in order to find gaps and criticalities. A second objective is understanding to what extent the emergent MOOC research reproduces the methodological problems observed more generally in eLearning and educational research.

The first step of the work consisted in defining the structure of a database destined to host the relevant information about the papers analyzed. The database records were structured as reported in Table 1. Each record corresponded to a paper and contained “objective” information like title, author(s), keywords, etc; followed by a characterization of the methodological approach adopted. In order to systematize the process, the concept of “methodological approach” was operationalized on the bases of the educational research literature (Arthur, Waring, Coe, & Hedges, 2012; Gorard, 2004; McKenney & Reeves, 2013; Trifonas, 2009) by taking into consideration:

a) the research aims, 9 categories were identified with a bottom up approach;

b) the research paradigm, following a rather traditional classification that distinguishes between quantitative, qualitative, mixed-methods, design-based research, literature review, theoretical-conceptual research;

c) the methods used for data collection, including classical methods (like surveys) but also newer ones like Tracking data for Learning Analytics (LA);

d) the methods used for data analysis, including classical ones (like statistics) and more recent ones (like Data visualization within LA).

e) type of research, distinguishing between desk-research, descriptive research and interventions.

Each dimension described above corresponds to one of the descriptive fields of the database records, with the exception of dimension that corresponds to two fields, one for data collection and one for data analysis. The criteria adopted to classify the articles were agreed upon by the researchers involved in the study.

<table>
<thead>
<tr>
<th>Data Field</th>
<th>Data Sub-field</th>
<th>Assigned values</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-Article Identification</td>
<td>Article Title</td>
<td>Title as published</td>
</tr>
<tr>
<td></td>
<td>Type of article</td>
<td>Range of values: Journal article, conference proceedings, reports, magazine articles. For this review, only journal articles were considered.</td>
</tr>
<tr>
<td>Source</td>
<td>Journal title</td>
<td></td>
</tr>
<tr>
<td>Author/s</td>
<td>Author(s) name and surname</td>
<td></td>
</tr>
<tr>
<td>Publication date</td>
<td>Month/Year</td>
<td></td>
</tr>
<tr>
<td>Key words</td>
<td>Keywords as published</td>
<td></td>
</tr>
<tr>
<td>Abstract</td>
<td>Full abstract as published</td>
<td></td>
</tr>
<tr>
<td>2-Research aim</td>
<td>The guiding research purpose: e.g. to produce a literature review or to study aspects of learning processes in MOOCs.</td>
<td></td>
</tr>
</tbody>
</table>
The second step consisted in selecting the literature to be examined. The articles included in the sample were published from January 2008 to May 2014 in peer-reviewed journals. Four specialized databases, namely SCOPUS, Web of Science (WOS), Educational Resources Information Centre (ERIC) and Directory of Open Access Journals (DOAJ), were queried searching for the term “MOOC” in the title, the abstract or keywords, both as acronym and as complete phrase. This search generated more than 300 papers, which were filtered by eliminating: a) duplicated papers, due to overlaps between databases, b) papers with full text in languages other than English, c) pieces of work other than research papers (reports, position papers, magazine articles, etc). An additional filter was applied to the resulting corpus of papers: proceedings papers were excluded from the analysis for this article, but they were included in the database of reviewed literature for a further phase of analysis. This process led to a sample of 50 journal papers indexed by at least one of the above databases.

To this sample, the papers published in a special issue on MOOCs of the “eLearning papers” journal were added. These ten additional articles were not indexed in the four databases consulted but were often cited by the other papers. All in all, 60 journal articles were analyzed. Table 2 shows the journals consulted and the distribution of papers per journal.

<table>
<thead>
<tr>
<th>Journal</th>
<th>No. Of articles</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>3. American Journal of Distance Education</td>
<td>1</td>
<td>(Zutshi, O’Hare, &amp; Rodafinos, 2013)</td>
</tr>
<tr>
<td>4. Australian Academic &amp; Research Libraries</td>
<td>1</td>
<td>(Barnes, 2013)</td>
</tr>
<tr>
<td></td>
<td>Journal Name</td>
<td>Volume (Year)</td>
</tr>
<tr>
<td>---</td>
<td>--------------------------------------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>6.</td>
<td>Canadian Journal of Learning and Technology</td>
<td>1</td>
</tr>
<tr>
<td>7.</td>
<td>College &amp; Research Libraries</td>
<td>1</td>
</tr>
<tr>
<td>8.</td>
<td>Distance Education</td>
<td>2</td>
</tr>
<tr>
<td>9.</td>
<td>Educational Research</td>
<td>1</td>
</tr>
<tr>
<td>10.</td>
<td>eLearning Papers</td>
<td>10</td>
</tr>
<tr>
<td>11.</td>
<td>Electronic Journal of e-Learning</td>
<td>1</td>
</tr>
<tr>
<td>12.</td>
<td>European Journal of Open, Distance and eLearning</td>
<td>4</td>
</tr>
<tr>
<td>13.</td>
<td>First Monday</td>
<td>1</td>
</tr>
<tr>
<td>14.</td>
<td>Intelligent Systems, IEEE</td>
<td>1</td>
</tr>
<tr>
<td>15.</td>
<td>International Journal of Emerging Technologies in eLearning</td>
<td>1</td>
</tr>
<tr>
<td>16.</td>
<td>International Journal of Self-Directed Learning</td>
<td>1</td>
</tr>
<tr>
<td>17.</td>
<td>Journal for Advancement of Marketing Education</td>
<td>1</td>
</tr>
<tr>
<td>18.</td>
<td>Journal of Asynchronous Learning Network</td>
<td>1</td>
</tr>
<tr>
<td>19.</td>
<td>Journal of Computer Assisted Learning</td>
<td>1</td>
</tr>
<tr>
<td>20.</td>
<td>Journal of Educational and Social Research</td>
<td>1</td>
</tr>
<tr>
<td>21.</td>
<td>Journal of Online Learning and Teaching</td>
<td>11</td>
</tr>
<tr>
<td>22.</td>
<td>Journal of Universal Computer Science</td>
<td>2</td>
</tr>
<tr>
<td>23.</td>
<td>Journal ACM Transactions on Computer-Human Interaction</td>
<td>1</td>
</tr>
<tr>
<td>24.</td>
<td>Learning, Media and Technology</td>
<td>1</td>
</tr>
<tr>
<td>25.</td>
<td>Open Learning: The Journal of Open, Distance and e-Learning</td>
<td>1</td>
</tr>
<tr>
<td>26.</td>
<td>Open Praxis</td>
<td>1</td>
</tr>
<tr>
<td>27.</td>
<td>Research in Learning Technology</td>
<td>1</td>
</tr>
<tr>
<td>28.</td>
<td>Revista de Educació a Distancia</td>
<td>1</td>
</tr>
<tr>
<td>29.</td>
<td>Turkish Online Journal of Distance Education</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>60</td>
</tr>
</tbody>
</table>

**Table 2 – Distribution of the papers in the Journals consulted**

The third step of our work was the analysis of the papers collected. At first, the article’s identification data (field 1 of Table 1) were introduced in the database. Afterwards, one of the researchers read and filled in all the database records with the more subjective information about the articles (fields 2 to 5 of Table 1). Finally, another researcher rated five randomly chosen articles in order to double check the initial rating results. Across the subjective fields, a
92% rate of agreement was reached: the percentage was calculated by dividing the number of agreements per field by the total number of items evaluated by both researchers. Disagreements were then resolved through discussion involving all the researchers.

4. Results

This section presents the results of the data analysis. For every dimension, a brief explanation is provided; moreover, descriptive statistics about the data, data interpretation and some insights on the qualitative data examined (key words and article’s abstracts as well as the whole article’s text) are also given.

4.1. Dimension one: Research Aims

In order to identify the research aims of each article, the researchers analyzed the information provided in each paper about the problem(s) and the purpose of the research work and subsequently identified the research aims as an overarching category based on this information. For example, where the problem concerned how to design a culturally inclusive MOOC, and the purpose was the investigation of cultural translations of MOOC resources, “Design for Learning in MOOCs” was chosen as a research aim. This research aim covers all the papers aiming to find answers to design problems in MOOCs. The nine research aims identified for the 60 papers examined are shown in Figure 1. While the titles of most of these research aims can be considered self-explanatory, the distinction between “Design for Learning” and “Teaching processes in MOOCs” deserves an explanation, as it lies in the fact that the former refers to the representational tools and strategies regarding MOOC design, while the latter has to do with the more pragmatic teaching strategies and activities used within MOOCs. The “MOOCs’ pedagogy” research aim, instead, includes papers that deal with pedagogical issues (both theoretical and conceptual) of the two above-mentioned categories (“Design for Learning” and “Teaching processes in MOOCs”).

![Figure 1 – Distribution of the papers in the Research Aims dimension](image)

Figure 1 shows that 18 papers aim to investigate LPM (Learning Processes in MOOCs, 18/60), followed at a distance by DLM (Design for Learning in MOOCs, 10/60), MP (MOOCs’ Pedagogy, 8/60) MID (MOOCs for Institutional Development, 7/60), CMET (Contribution of MOOCs to Educational Theory, 7/60) and MID (MOOCs for Institutional Development, 7/60).
Within the 18 LPM papers, it is possible to identify different sub-categories: the most important one regards the problems of completion rates in MOOCs, followed by factors hindering or promoting learners’ engagement; then by the issues of self-directed learning and literacies required to participate in MOOCs, including the proficient use of digital technologies and social media. Furthermore, learners’ feelings when taking a MOOC is one of the aspects considered in several works within the LPM category. These data allow us to conclude that the major research areas within the field are connected with the way MOOCs are designed and their impact on formal and informal learning processes; this focus could be interpreted as a search for the MOOCs’ identity as educational phenomenon, strengthening their status amongst other more consolidated fields of eLearning research. It should also be noted that there were only four studies investigating TPM (Teaching Processes in MOOCs, 4/60), which is in line with the idea that learning processes and design issues are by and large the kernel of the debate, while the role of instructors/facilitators is considered less important. Only three studies focused on MASM (Methodological Approaches to Study MOOCs, 3/60), which highlights that there is little concern about the problem of research methods and their connections with both practical and theoretical issues to be faced. Another poorly investigated dimension is the one devoted to the description and testing of Technological tools for MOOCs (TM, 1/60), even if the database search performed to find the papers included “Computer Science” and “Engineering” among the subject areas. Regarding Literature Reviews (LR, 2/60), the small number of papers in this category is understandable given the relatively young field of research.

4.2. Dimension two: Research paradigm

During the ‘80s a strong debate, engaging social science researchers mainly from the education field, highlighted two already existing “worldviews”, with their implications for thinking, designing and conducting research: the quantitative and the qualitative methodological “paradigms” (Lincoln, Lynham, & Guba, 2011). They were ascribed to diverse ontological and epistemological visions about research and, as such, many regarded them as incommensurable (Gage, 1989). While the quantitative tradition was objectivist and nomothetic (in search of explanation/predictions), the qualitative tradition was interpretivist and ideographic (in search of deep understanding) (Lincoln et al., 2011). More recently, the debate regarding methodologies for research effectiveness and quality has led to the configuration of a third field (Gorard, 2004), the so-called “Mixed Methods”, regarding the possibility of mixing methods coming from quantitative and qualitative traditions in several phases or research activities, with a pragmatic approach, aiming to find out what works (Tashakkori & Teddlie, 2014). During the ‘90s another line in educational methodologies emerged, namely “Design Based Research” or DBR (Brown, 1992; The Design-Based Research Collective, 2003); according to Anderson & Shattuck (2012), DBR is to be considered an alternative to quantitative, qualitative or mixed methods research, since it entails the design of an intervention aiming to change an existing situation in a natural setting. In our study, in addition to the above-mentioned categories, several papers did not refer specifically to one of the above paradigms, in that they either propose conceptual schemes or models created by the authors or discuss theoretical aspects and frameworks for analysis: we labelled these papers as “theoretical-conceptual”. Figure 2 shows the distribution of the papers examined for this dimension, including a category “not clear” for papers that were not ascribable to any of the identified categories.
A first analysis of our sample showed that in general, there was little concern about explicitly declaring the research paradigms embraced. Several articles explained the research method adopted without taking a position with regard to their “worldview”. Consequently, articles were assigned to a category within the dimension “methodological paradigm” based, in the first place, on the authors’ claims, and in the second, on our deductions after their analysis.

In this dimension, the largest category (18/60) consists of theoretical and conceptual studies. This is not surprising, due to the fact that MOOCs are a recent phenomenon. Theoretical papers focused on issues like how MOOCs relate to Connectivist Learning or to Collaborative Learning. Conceptual papers consisted of categorizations of the existing literature, creation of dimensions/concepts and their application to a case or a classification of cases using a proposed framework. It is to be noted that the concepts were mostly created by the authors of the papers in this category, based on the literature or on their own experience, not on the basis of a systematic method, such as literature reviews, grounded theory or statistical analysis. These types of studies were conspicuously present among those classified as CMET, MP and DLM regarding their research aims.

The second category in terms of frequency was that of “Mixed Methods” (12/60). While the complexity of phenomena under analysis in educational research makes this approach particularly useful (Gorard & Cook, 2007), it should be highlighted that many studies within this category did not take a clear position, but they described the method used and hence introduced a profusion of data that was not always coherent with the main research question or the purpose of the study. A clearer position would encompass, instead, mentioning the approach, explaining how the research problem was formulated and how, when and why the quantitative and qualitative data were collected. This was even more evident for the studies that we classified as “not clear” (9/60): this is the case of papers where the methodological approach was not plainly specified, rather, there was an initial formulation of the research problem followed by the analysis of several types of data (obtained through LA, surveys results, learning artifacts produced by the students, excerpts of discourse taken from online fora or social network analysis results). This typology of studies was frequent among those classified as LPM and TPM research aims.

The studies classified as “qualitative” (7/60) and “quantitative” (9/60) declared the methodological approach adopted more frequently; overall, the authors showed awareness...
Regarding their “worldview”. With regard to design-based research studies, only three studies (out of the 5 that were classified as such) openly declared this approach, with the others referring to “interventions” or “pedagogical experiments”.

### 4.3. Dimension three: Methods

The third dimension analyzed, “Methods” includes both methods for “data collection” and methods for “data analysis”. This type of classification is schematic and there was an intense debate within the research team regarding the feasibility of the classification, especially for theoretical papers. In many cases, the articles introduced clear statements and information regarding methods; in others, we defined the methods taking into account the information provided in the papers, on one hand, and the literature on research methodology, on the other. As for the methods adopted for Data Collection (Figure 3), we observe that the most frequent category (22/60) is the one we called “Conceptualisation of dimensions”, consisting of the description/classification of different MOOC characteristics, usually aiming to support a thesis or substantiate a claim of the authors (i.e. exploring how teaching, learning, assessment, and pedagogy can be characterized in a cMOOC versus an xMOOC). The second largest group is “Survey” (18/60); many researchers adopted questionnaires - pre and post training - to study the learners’ profiles, their activities and feelings during and after their participation in MOOCs, and their opinion about pedagogical practices. The surveys mostly consisted of a set of closed, multiple-choice questions but occasionally also included open-ended answers, which is aligned with a “Mixed Methods” research paradigm (§ 4.2).

Little use was observed of other, more specific and systematic methods aligned with the qualitative paradigm (interviews, virtual ethnography) or the quantitative tradition (quasi or controlled experiment).

![Figure 3 – Distribution of the papers according to the Methods adopted for Data Collection](image-url)
As for the methods adopted for data analysis (see Fig. 4), qualitative studies hardly ever specified the data collection and analysis methods. Very few of these studies explicitly stated that discourse or thematic analysis was used; among these, in some cases, the data and the methods of analysis appeared to us as not consistent, due to the lack of excerpts of discourse with corresponding analysis or semantic categories used to analyze the discourse. Still with regard to qualitative methods, in at least two studies the authors mentioned “virtual or online ethnography” as the method chosen. This type of analysis should encompass the presentation of cultural artifacts and description of groups’ rituals and heroes, as the deep exploration of a cultural group would require. However, we were not able to find any evidence of the actual implementation of this type of analysis.

For the quantitative case, the massive use of descriptive statistics sharply contrasts with the sporadic use of other types of statistical instruments like correlation, factor analysis and inferential statistics. This can be interpreted as a consequence of the type of studies - exploratory, rather than experimental - which is coherent with the fact that MOOCs are a nascent field of research. However, a tendency to “describe cases” or “generate frameworks” to conceptualize the problem, even if appropriate for an initial phase of development of a research field, should be followed by the use of more systematic approaches, such as design experiments or controlled experiments. It is to be noted that only two out of the 60 studies analyzed implemented randomized experiments (see §4.4)

Furthermore, we observed an interest in using innovative quantitative tools for data collection and analysis. Thirteen studies adopted Learning Analytics (LA) (Ferguson, 2012) for data collection. To do so, they tracked data through the Learning Management System (LMS) used or the social networks integrated into the course’s activity. These data mostly regard the online activities (clicks on specific resources, records of interactions, etc.). LA techniques are used in early works as well as in the more recent cases of scholarly literature on MOOCs. Eight studies tend to use LA as a method to provide information regarding the frequency of interventions or interactions amongst participants. However, in these papers, the connections between the constructs discussed were loose and the data were not elaborated further than descriptive statistics. Few papers adopt LA in more advanced ways to show learning patterns or generate predictive models. One paper (Bieke et al. 2014) addresses the way LA can be connected with specific theoretical constructs (e.g. Networked Learning) and their operational implementation through a specific method (Social Learning Analytics) and related tools (Social Learning

![Figure 4 – Distribution of the papers according to the Methods adopted for Data Analysis](image-url)
Analytics Tool). Four papers propose mathematical models to explain relationships between the variables involved and to predict future learning behaviors on these bases.

4.4. Dimension four: Research Types

One of the initial discussions within our research team had to do with “the way the research field on MOOCs is developing”, and how it is connected with the “full cycle of educational research” framework (Gorard, 2004; Gorard & Cook, 2007; Raffagellli & Persico, 2014). Very briefly, Gorard’s claim is that educational research in a given field evolves along a cycle covering seven phases of development, from the acknowledgement of a problem’s existence to the studies aimed at proposing solutions and analysing impact or achieving technology transfer. As Gorard explains:

“In this cycle, reviews and secondary analyses might appear in Phase 1, theory building and small-scale field-work in Phase 2, etc., with smaller experiments being part of Phase 5 and full randomized controlled trial only appearing once in Phase 6” (Gorard, op.cit, p. 319)

We classified the research types of the 60 articles considered in this study and referred to Gorard’s framework to match them with the sequence of phases he identified. This would provide further evidence to understand the state of the art of research in MOOCs, showing at which “phase” of the full cycle of educational research this field of research is. To operationalize Gorard’s phases we defined the dimension “types of research” by allocating the studies to the following categories:

- Desk-research, - aligning with phase 1 - including theoretical and conceptual research as well as literature reviews;
- “Descriptive” - aligning with phase 2 -, which was further divided into “descriptive-exploratory” and “descriptive-observational case studies”. Both categories involve studies aimed at exploring phenomena without intervening in an existing situation, with the former focussing on more than one case and the latter concentrating on a single case;
- “Interventionist” - aligning with phases 3-7 - which was further divided into
  - “Interventionist-case study”, focused on a unique case study, but implementing an intervention to analyze educational processes and impacts;
  - “Interventionist-design experiment”, design experiments (in real educational contexts) promoting cycles of trialing accompanied by small improvements and further analysis (intra-subject analysis);
  - “Interventionist-random experiment”, randomized experiments (inter-subject analysis).

Figure 5 shows the yearly distribution of the articles analyzed according to the “type of research” dimension, with a peak of publications in 2013 and the first half of 2014. In this picture, the distribution of types of studies is not in full accordance with the phases of Gorard’s cycle. There are different types of studies coexisting along the considered period: interventionist studies alternate with descriptive ones; in 2013 there is an impressive production of desk-research; in 2014 (first half) there is a high number of descriptive studies. This could be explained by the “compression” of studies in the last two years, but it also suggests that this field is not yet steadily evolving towards more mature stages.

Figure 5 – Distribution of the papers per year and “Research Type”
According to Table 3, showing the same data as Figure 5 without the temporal perspective, the desk research studies together with the descriptive studies (based on both qualitative or quantitative approaches) are the vast majority. In the light of Gorard’s cycle of the “evolution” of a research field, this further confirms that research on MOOCs is in its early stages devoted to characterizing the object of study, conceptualizing the phenomena, and identifying more clearly and systematically the main research problems.

5. Discussion

The evidence reported in the section above shows an increasingly active evolution of research on MOOCs, which became rather hectic during the last two years considered. In the following, we discuss the results of the data analysis.

5.1. Research aims

The panorama of research aims shows that interest has been primarily concentrated on the study of learning dynamics in MOOCs, the way MOOCs can be designed, their relationship with...
educational theory as well as their impact on the present and future of higher education institutions. Teaching processes are hardly under the lens of researchers and this can perhaps be explained because massive courses generally rely on a high degree of self-regulation of learners and their interactions with peers with different degrees of expertise, rather than traditional teaching. On the contrary, the small number of studies concentrating on technological tools suitable for open and massive online courses is, in our opinion, quite striking. The availability of tools able to support massive educational interventions is a pressing need for future research, since many traditional LMSs are likely to turn out to be inadequate and need to be enhanced with or replaced by robust tools supporting social networking, learning analytics, audio/video broadcasting, self regulated learning, etc. The lack of these studies has several possible explanations. It is possible that only a limited number of MOOC designers have dared to experiment with innovative tools, because technology failures in massive courses are a bigger institutional risk than they are in courses with a more manageable audience size. Another possible explanation is the fragmentation of the disciplinary contributions to MOOC research, with much work in the area of Computer Science being published with different keywords (e.g. cloudcomputing). However, the exploration of the potential of these tools (Learning Analytics, data visualisation and Educational Data Mining, to mention only some) is not just a matter for computer scientists, as it requires a close connection between definition of the pedagogical problem, development of a prototype solution, and tailoring of the mathematical and statistical procedures to extract and suitably represent significant data. In other words, what is needed to produce high impact results is a joint effort of several experts. Finally, while the limited number of review papers can be explained by the fact that this field of research is still rather new, the lack of papers addressing the methodologies for research in this area suggests that, at least so far, researchers have felt little need to discuss whether new methods are needed to investigate this phenomenon or whether traditional methods should be adapted in some way to be usable in this area. Actually, we would not be surprised if, in the near future, there was a burst of interest in this area, possibly in conjunction with the investigation of tools, such as LA tools, that promise to provide a wealth of data difficult to integrate with each other and to analyze with traditional methods.

5.2. Research Paradigms

The panorama of research paradigms adopted in the papers analyzed shows that, although data-driven approaches (design based, qualitative, quantitative or mixed method studies all together) outnumber conceptual and theoretical papers, the latter still form the biggest category. This is possibly due to the fact that researchers in the field of MOOCs are making an effort to define the main constructs needed to drive empirical research. However, it might also be showing an “inflated” debate poorly underpinned by empirical evidence. The number of papers where we did not manage to identify the worldview of the authors is rather high and confirms the lack of attention that researchers are paying to the methodological aspects involved in this field of research. No new research paradigm, specific for this research field, has emerged so far.

5.3. Research Methods

As for research methods, and in particular data collection methods, it strikes the attention that observation and critical analysis (termed “Conceptualisation of dimensions”) is by far the most frequently used method, confirming that the area is in a rather preliminary phase of study where hypothesis building is still one of the main activities. Surveys are also frequently used, which is understandable because of the large number of people involved that can provide useful information, even if it is likely that most of the data collected come from those who have
completed the courses while, perhaps, those who dropped out might provide even more useful information. In many cases, anyway, the surveys were filled in by a small percentage of participants, therefore revealing that this data collection tool is an insufficient source of information in the context of massive courses.

Special attention is deserved by LA techniques that integrate tools for data-collection as well as data-analysis, showing that the border between the two is becoming more and more blurred. We note, however, that most research analyzed hereby did not unfold the whole potential of LA since the data tracked were visualized in terms of descriptive statistics to “show” trends of participation, while more refined analysis for modelling and predicting learning processes was hardly used. Other, more time consuming methods, like focus groups or interviews, are likely to be of little use with the large cohorts of people participating in MOOCs, but may still be used with MOOC designers or facilitators.

Finally, among data collection methods, it appears that learning assessment and self-assessment do not seem to be a primary source of information used in research in MOOCs, and coherently, data analysis does not devote much attention to learning outcomes. However, we deem it essential that research in this area takes into consideration the outcomes, in terms of learning, of this new type of courses, as well as the methods for measuring such outcomes as part of more complex learning pathways within lifelong learning. One promising approach, although still to be fully explored, are Digital Badges (Moore, 2013; Cucchiara, Giglio, Persico & Raffaghelli, 2014).

In data analysis, descriptive statistics is used much more frequently than other statistical methods. This is in line with the fact that this research area is still rather young and, as such, more involved in understanding the general scenario rather than finding out the answers to causal questions, although these are of major importance for educational research and policy. More generally, it is not surprising that quantitative methods are privileged in a field where more time consuming methods of analysis are not much viable due to the abundance of data sources, and indeed LA studies are already producing interesting results in terms of data visualisation, thus providing, literally, a better picture than tables full of numeric data.

5.4. Research Types

With regard to research types, our results roughly align with Gorard’s point (Gorard, 2002): “We also agree that experiments are currently lacking in education research practice writ large, and that most education research gets stuck in Phases 1 to 4. In other words, it is stuck working towards a randomized trial that hardly ever gets done” (Ibidem). This is particularly true of MOOCs research, whose evolution so far has been short and uneven.

It is valuable, in any case, that a high number of papers are going beyond conceptual descriptions to interventionist approaches, however small scale, where there is an effort to prototype, trial and analyze impact. This is perhaps suggesting that, in the near future, we will possibly witness research activities regarding several cycles of analysis of impact or experimental research. However, we wish to highlight that we do not consider Gorard’s “phases” in a hierarchical or linear way. A robust methodology encompasses good research questions, addressing relevant educational problems, and implementing the type of study that gives better answers to those questions. This means that an exploratory study attempting to configure a research field and clarify research problems is as important as a randomized experiment.

5.5. Further remarks
In this section we provide some considerations which are not directly supported by the data collected in our database, but have emerged from the qualitative analysis of the papers and have been object of discussion among the researchers involved in this study. Firstly, we agree with Ebben & Murphy (2013) who identified two periods in MOOC research: a first phase of enthusiasm and experimentation followed by a second one, more critical and objective, oriented towards the exploration of strategies to improve access and innovate higher education.

Secondly, we noted a lack of debate about methodological issues in this research area, in spite of its peculiar features that might require bespoke methods, apt to collect, analyze and draw conclusions from the profusion of data which emerge from this kind of course. Lastly, in some of the papers analyzed, we noted that the data produced by a MOOC seemed to determine “ex post” the methodological approach rather than the opposite. This problem has already been pointed out by Ann. L. Brown (1992) with regard to educational research in general: many real-life experiments produce a maze of data, among which it is hard, but necessary, to select those that answer the researcher’s question. In addition, many of the studies present quantitative data such as the number of blog- and forum-posts produced by their course participants or the number of social network interactions that took place, often in the form of descriptive statistics. However, the same studies do not provide reference parameters to support the interpretation of their data, as if the sole number would be sufficient to “speak”.

6. Conclusions, limitations and further research directions

This literature review focuses on the methodological approaches of the papers analyzed, defined in terms of research aims, research paradigms embraced, methods of data collection and analysis, and type of research. Several issues have been faced during the study, such as defining the classification categories, achieving a reasonable degree of objectivity in classifying the studies, choosing a framework to analyze the trends and identifying the critical aspects of MOOC research taking into consideration the evolving methodological debate concerning educational research at large.

The results show that this field of study, so far, has focused on learning processes and pedagogical issues, impact of MOOCs on pedagogical theory and on educational institutions (especially in higher education), with many studies based on desk research and a lower number of experimental investigations. Experimental studies, in any case, are usually small scale ones, often based on a case study of a single MOOC. Relatively little interest in tools and methodological aspects has emerged, which is likely to need to be reconsidered in the light of the unprecedented problems faced. Traditional approaches to data collection and analysis prevail, along with some relatively new ones, such as LA, with emphasis on the learning processes (what did participants do?) rather than the learning outcomes (what did participants learn?). The research methods adopted tend to be preeminently theoretical and sometimes empirical, but in this case they are not very sophisticated, because they are oriented towards understanding the phenomenon and building hypothesis rather than answering specific research questions. In the framework of Gorard’s full cycle of educational research, research on MOOCs is mostly collocated at the initial phases of development, although it seems to progress in an uneven way, with a burst of interest in 2013 and 2014. The evolution in these two years was rather hectic and scarcely attentive to methodological aspects. This is a weakness that this area of research shares with the wider panorama of educational research but, in our view, research on MOOCs faces some specific needs in terms of tools, methods and aims, that require researchers to engage in a methodological debate and, possibly, to propose some innovative solutions to handle the massive and open features of these courses.
The limitations of this study have to do with the fact that this research area is evolving very fast and the literature on MOOCs is extremely rich. We chose to analyze journal papers only, as in many literature reviews, but it is well known that the debate about this phenomenon also takes place at conferences and in their proceedings, in blogs and social networks, as well as in non-scientific magazines. Our choice was driven by the need to consider only fully-fledged research, rather than opinions, work in progress and non-substantiated claims. However, we are aware that we might have missed some significant contribution to the debate. Besides, while preparation of this paper is in progress, new analysis on research trends in the field of MOOC are being published on the subject.

There are several directions that could be taken to advance research in this area. One example is the investigation of the relationship between the research aims and the methods adopted - particularly those regarding the treatment of big data. Another important topic has to do with the way ethics in the use of public data is dealt with, in order to identify criteria and approaches to handle informed consent (some institutions have clear policies in this regard, but others do not).

Thirdly, the geographical areas originating the papers could be investigated, in order to understand whether there are cultural differences or similarities as to the concerns, the methods used and the results obtained.

Last but not least, we believe there is a need for a sort of “methodological tuning” of educational research methods to the field of research on MOOCs, a need determined by the complex and new nature of the educational problems under analysis in this area. The different backgrounds and skills of the researchers involved in eLearning (Conole & Oliver, 2006) lead to different epistemological and ontological conceptions about the nature of the issues faced and the way these should be studied. This is, at the same time, a strength and a weakness. It is a strength, because it allows us to better face a phenomenon (MOOCs) which is multifaceted and multilevel, in that it requires the study of learning and teaching processes, but also the solution to policy issues such as impact on institutional and social development and the development of suitable technological environments. It is a weakness, because it produces results that are sometimes difficult to integrate into a coherent picture. The “methodological tuning” we advocate should result in the production a more mature and extensive application of systematic research methods to the field of MOOCs.

We are aware that every effort to “catch Proteus” can be tricky, since applying a scheme to a changing flux can end up in imposing a scheme that might mislead rather than guide the creation of knowledge. In the Greek myth, Proteus is the owner of full wisdom. Such a condition does not exist: in line with critical realism (Archer, Bahaskar, Collier, Lawson and Norrie, 1998), in social sciences we could talk about perspectives adopted by every scientific community, politically and culturally contextualized. What we have tried to maintain in this paper is that the mechanisms to become aware of “how” scientific research is being undertaken cannot be missed in every evolving field. This is, to us, “catching Proteus”.

7. References


Rodriguez, O. (2013). The concept of openness behind c and x-MOOCs (Massive Open Online Courses). *Open Praxis, 5*(1), 67–73. doi:10.5944/openpraxis.5.1.42


