Bibliometric Analysis of Technology-Supported Language Learning: LMOOC Trend in China

AbSTRACT

Interest in LMOOCs (Language Massive Open Online Courses) has increased in terms of research and number of courses offered, due to their potential for developing training processes and for working on these contents with technologies. Situating the interest in the publications made on LMOOC in China, this study aims to bibliometrically analyze said scientific production in Scopus. The final sample was made up of 134 publications, which were analyzed using different bibliometric techniques (bibliographic coupling, co-citation, co-occurrence). The results indicate a stabilized scientific production since its inception in 2014, with further development in the years of the COVID-19 pandemic. Indexing predominates in the areas of computer science and social sciences. Contributions to conferences and articles predominate. The most cited publications focus on systematic reviews, student autonomy and active methodologies combined with LMOOC. Other influential publications focus on defining the quality criteria to carry out an LMOOC and the emotional factor involved in the development of these courses. Among the potential future courses of action, highlight the adoption of more personalized formats such as the SPOC or the inclusion of methodologies such as the flipped classroom. It is concluded that this study on scientific production will allow the development of new avenues of research on LMOOCs from what already exists.

KEYWORDS LMOOC; Language learning; Research methods; Information and Communication Technologies (ICT).
1. INTRODUCTION

Thanks to the advances in Information and Communication Technologies (ICT) and Internet access in the last two decades, the use of open educational resources has seen a proliferation in teaching and learning processes (Bethencourt-Aguilar et al., 2021; Comas-Quinn, & Borthwick, 2015), especially in language teaching (Martín-Monje, & Borthwick, 2021) where MOOCs (Massive Open Online Courses) for foreign language learning (LMOOCs) represent one of the key elements.

LMOOCs are defined as massive open online courses specifically designed to teach languages (Hsu, 2021). In particular, this can be understood as “dedicated web-based online courses for second languages with unrestricted access and potentially unlimited participation” (Barcena, & Martin-Monje, 2014, p. 1). LMOOCs have flourished since 2013, providing foreign language learners with a wide range of learning resources, increasing their linguistic input and connecting students from different countries and regions with diverse language exchange opportunities (Jitpaisarnwattana et al., 2021; Sallam et al., 2022). Typically, these courses are available online and are either free (Luo, & Ye, 2021) or have a basic free option with paid options for additional content or certificates (Caro-Barek, 2022). LMOOCs offer a wide variety of languages and are usually designed to provide students with skills in reading, writing, listening, and speaking in the selected language (Lebedeva, 2021).

LMOOCs often use online learning platforms that offer videos, interactive exercises or even exams (Gharawi et al., 2020; Ruiz-Palmero et al., 2021). Some of these courses also feature real-time interactions, such as live chat sessions with instructors or tutors (Cinganotto, & Cuccurullo, 2019). Moreover, these types of courses can go beyond formal instruction. As stated by Vázquez-Cano et al. (2018, p. 179) these courses can be “interesting formative modalities to develop in this polychromatic plurality of formal and informal educational contexts. In this context, Gil (2021) states that LMOOCs help all those individuals who want to carry out non-formal education or even those who find themselves in vulnerable situations such as professional and social exclusion. With this, it would be possible to meet the linguistic needs of any type of student body (Negre et al., 2018).

In China, where this study is contextualized, the use of MOOCs or in our case, LMOOCs has been generally welcomed by the educational community as a revolution which can help democratize education and...
promote educational equality among students (Peng, & Jiang, 2022), especially with the rapid development of internet usage and low-cost mobile technologies (Law et al., 2019). Within this context and bearing in mind the exponential number of LMOOCS existing today, this need has arisen to conduct a critical, systemic and detailed review of existing recent research which will help facilitate a deeper understanding and knowledge about this type of courses in China. With this new analysis it will be possible to identify gaps in the scientific literature on LMOOCS in China and suggest new directions and challenges. This review paper aims to contribute to this research.

Taking all of the above into consideration, the focus will be on those publications on LMOOCs in which some of the researchers have their professional affiliation in a Chinese research institution. The aim of this study was to analyze bibliometrically the scientific production on LMOOCs in China in the international Scopus database.

In relation to this, different research questions were established:

1. How has the scientific production of LMOOCs in China evolved with respect to the variables known as year of publication, areas of knowledge, type of document, institutional affiliation and relevance of publications?
2. What are the most influential publications in the topic of study?
3. What are the main lines of research related to LMOOCs in China?

2. MATERIAL AND METHOD

The reality of LMOOC research in China was analyzed through a bibliometric study. As a research methodology, it is a meta-analysis technique that explores scientific production with respect to the object of study (González et al., 2020). During its application, criteria are established to analyze the progress and development of publications from a qualitative and quantitative perspective in a thematic area, considering different variables. This technique is supported by multiple studies that have developed and implemented it in recent times (Colomo et al., 2022; 2023; Marín et al., 2021; Mielgo et al., 2022).

The selection of documents was made in the Scopus database. Scopus was chosen because it has a large, rigorous and high-quality scientific production, including multiple areas of knowledge within the academic field (Cívico et al., 2022). The keywords and Booleans used to choose the sample (search command) were “MOOC” OR “MOOC’s” OR “LMOOC” OR “LMOOC’S” AND “Language”. The command was applied within the title, keywords and abstracts. The search returned 890 documents as of August 1, 2023, including all types of documents (articles, book chapters or papers, among others).

Screening criteria were applied to the sample of 890 documents (Figure 1), linked to the purpose of the study, using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement as a model. The first action was to restrict the results by country (China), so that any of the authors of the paper that investigated MOOCs and their application to language learning had their affiliation in an institution in that country. This brought down the sample size to 139 papers. Following this, research whose object of study was not directly related to LMOOCs was purged, eliminating another five papers. The final sample,
after applying the above limitations, consisted of 134 documents (132 in English, one in Chinese and one in Spanish). This sample, after being exported from Scopus in .csv format (comma separated values), was analyzed bibliometrically.

**FIGURE 1.** Flowchart of the study selection process based on the PRISMA statement.

To examine the sample, several bibliometric techniques were used: a) an analysis of the scientific production, evaluating how the publications have evolved considering certain variables; b) a bibliographic linkage, which allows us to know the references in common between publications, being able to determine the influence of the most relevant sources; c) a co-citation analysis, which allows us to know the number of times different articles are cited together; d) a word set analysis, where the most frequent descriptors/keywords indicate the main topics that have been worked on in the publications. The exploration of the relationships and links developed between publications was assessed using the VOSviewer software. This program allows visual representation of the relational nodes generated in each factor considered.

The variables that were considered in the analysis of the scientific production should be noted, as well as whether any inclusion/exclusion criteria were used to examine them. These criteria arise from the need to choose only those elements that are most representative, and not to reproduce all 134 documents in all the variables. Thus, in some variables, only the most frequent descriptors are selected because they are more relevant to the variable under study.

- **Year of publication.** This variable indicates the temporal distribution of the publications. As an inclusion criterion, publications from the last 10 years (2014-2023) were considered.
- **Subject area.** This allows us to know to which field of knowledge the publications belong. The subject areas considered are those that at least registered 10 publications.
- **Type of publication.** This variable indicates which format of publications were those that inquired into the object of study. In this variable there was no exclusion criterion, including all the documents that make up the sample.
3. RESULTS

The results section is structured on the basis of the bibliometric techniques used to answer the research questions. We begin with the analysis of scientific production, followed by the analysis of bibliographic linkage, co-citation and co-occurrence.

3.1. Analysis of scientific production

The 134 documents were analyzed considering the different variables that were stipulated.

3.1.1. Year of publication

Based on the criterion of including studies from the last 10 years (2014-2023), the final sample (after applying the different filters) also had no publications prior to this. The results reveal an upward progression from the first-time tranche (2014-2016) to the second tranche (2017-2020), with the exception of 2019 which only recorded eight publications. Within the last tranche (2021-2023), we have results that follow the line of the second time tranche (11 publications in 2023) and a potential growth in the two years recording publications around the COVID-19 phenomenon, where technology-mediated learning (resources, methodologies, etc.), were the focus of interest of researchers due to the demands caused by periods of isolation during the pandemic.

![Figure 2. Documents published per year.](image)
3.1.2. Subject area

Only areas with at least 10 publications were considered in this variable. In addition, it should be borne in mind that the assignment of a publication to an area in Scopus is multi-classification. This means that the same publication can be associated with more than one subject area, so that the number of publications associated with areas will always be greater than the number of publications in the sample. This is reflected in Table 1, where the sum of only the areas that meet the criterion is greater than the 134 publications in the sample.

**TABLE 1.** Subject area.

<table>
<thead>
<tr>
<th>Area</th>
<th>Number of publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer Science</td>
<td>98</td>
</tr>
<tr>
<td>Social Sciences</td>
<td>51</td>
</tr>
<tr>
<td>Engineering</td>
<td>41</td>
</tr>
<tr>
<td>Mathematics</td>
<td>26</td>
</tr>
<tr>
<td>Decision Sciences</td>
<td>18</td>
</tr>
<tr>
<td>Arts and Humanities</td>
<td>11</td>
</tr>
</tbody>
</table>

Computer sciences was the subject area with the most publications (98), with social sciences in second position, with about half of the records (51). The variety of areas is noteworthy, due in part to the technological component. This is due to the fact that LMOOCs are designed and created from this area, with a high weight of engineering (41). However, the role of the area of mathematics stands out, as it is transversal to the technological component and is not the focus topic of the MOOCs, having a high presence in spite of this.

3.1.3. Type of publication

Another important aspect was to know what type of publications had dealt with the subject of the study. Based on the 134 documents in the sample and without applying any exclusion criteria, Table 2 shows the results.

**TABLE 2.** Type of publication.

<table>
<thead>
<tr>
<th>Type of documents</th>
<th>Type of publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conference Paper</td>
<td>77</td>
</tr>
<tr>
<td>Article</td>
<td>54</td>
</tr>
<tr>
<td>Book Chapter</td>
<td>2</td>
</tr>
<tr>
<td>Review</td>
<td>1</td>
</tr>
</tbody>
</table>

Contributions to conferences were in the majority, with more than half of the documents published (57.5%), followed closely by articles (40.3%). Book chapters (1.5%) and reviews (0.7%) had an anecdotal presence.

3.1.4. Institutional affiliation

For this variable, only those institutions with at least three publications on the topic of study were considered. Although the sample reached 14 documents, the most prolific institutions were Huazhong Normal University and Tsinghua University, both with five publications. Peking University had one less (four), while several institutions had three publications: Northeast Normal University, Guangdong University of Foreign Studies, Beihang University, Harbin Institute of Technology, Beijing Normal University, National University of Defense Technology China and Hangzhou Normal University.

3.1.5. Most relevant publications

The criterion for selecting the most relevant publications was based on the total number of citations achieved. In this case, publications that achieved 17 total citations or more were considered (Table 3).
Although the paper by Wang et al. (2018) is the most cited paper, it does not have the best average number of citations per year. This rate of citations per year is headed by the paper by Sallam et al. (2022), being the second with the most citations in total. It should be noted that only the first work is prior to 2022, since the other three that make up this variable of most relevant publications are from that year, having good citation rates per year that may cause them to change their position in future years within the classification by total citations. Regarding topics, there are two papers focused on systematic reviews (Fang et al., 2022; Sallam et al., 2022), another focused on the student body (Ding, & Shen, 2022) and another on the impact of the use of an active methodology in LMOOCs (Wang et al., 2018). Regarding systematic reviews, the most relevant (by citations) is that of Sallam et al. (2022), where research published from 2012 to 2018 on massive open online language courses (LMOOCs) is analyzed. The results highlight that Spain is the most prolific country, with the “Universidad Nacional de Educación a Distancia” leading the way. The main findings include the advantages that the didactic use of LMOOCs can bring to learning due to their suitability or the motivation they achieve in students. The review by Fang et al. (2022), covering 10 years of studies (2009-2018), focuses on reviewing the research on open online language learning. Among their conclusions, it should be noted that most of the studies were empirical, predominantly on reading and writing, with self-directed and blended learning processes, especially analyzing factors such as motivation, interest and satisfaction in LMOOC learners. The work of Ding and Shen (2022) places learner autonomy in LMOOC as the main focus, using interviews as a research instrument. It should be noted that learner autonomy was contingent on factors such as metacognitive strategies, motivational control and emotional control to regulate learning. Regarding the impact of active methodologies in LMOOCs, Wang et al. (2018), analyzed whether the MOOC-based Flipped Classroom improved the development of oral proficiency and learner progress. Using a control group and an experimental group, the results indicated that the students who took the LMOOC with the Flipped Classroom methodology had a 25% faster progress rate. In addition, these students significantly improved their oral proficiency, especially in speech fluency but less so in language complexity and accuracy.

3.2. Bibliographic coupling

Bibliographic linkage allows highlighting the influence of a publication within the scientific production analyzed, due to its similarity with other documents. To do so, we considered the number of references in common

<table>
<thead>
<tr>
<th>AUTOR</th>
<th>AÑO</th>
<th>TITLE</th>
<th>MAGAZINE</th>
<th>QUOTES</th>
<th>QUOTES PER YEAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wang, An, &amp; Wright</td>
<td>2018</td>
<td>Enhancing beginner learners’ oral proficiency in a flipped Chinese foreign language classroom</td>
<td>Computer Assisted Language Learning, 31(5-6), 490-521</td>
<td>57</td>
<td>9.5</td>
</tr>
<tr>
<td>Ding, &amp; Shen</td>
<td>2022</td>
<td>Delving into learner autonomy in an EFL MOOC in China: a case study</td>
<td>Computer Assisted Language Learning, 35(3), 247-269</td>
<td>18</td>
<td>9</td>
</tr>
<tr>
<td>Fang, Hwang, &amp; Chang</td>
<td>2022</td>
<td>Advancement and the foci of investigation of MOOCs and open online courses for language learning; a review of journal publications from 2009 to 2018</td>
<td>Interactive Learning Environments, 30(7), 1351-1369</td>
<td>17</td>
<td>8.5</td>
</tr>
</tbody>
</table>
among the documents in the sample, applying a backward citation chaining. With this we were able to find the referent publications, within the subject of LMOOCs in China. The bibliographic linkage analysis was developed with the documents as the unit of analysis, setting as inclusion criteria that the document had been cited a minimum of four times. The relationship nodes generated with the 10 items that met the criterion are shown in Figure 3. 4 sets of documents were formed by the coupling relationship between them.

It is worth highlighting the higher linkage intensity (total link strength 21) among the documents grouped around the blue cluster (Fang et al., 2022; Sallam et al., 2022), being the third and first documents with the most citations. It should be noted that both studies were systematic reviews of LMOOC research. The yellow cluster has a lower intensity (total link strength 9), although by citations between both papers (Ding, & Shen, 2022; Luo, & Ye, 2021) it would be in second position. Learner autonomy in language courses through MOOCs in China was the topic researched by Ding and Shen (2022). Luo and Ye (2022), who, on the other hand, worked on defining criteria from a learning perspective to ensure the quality of MOOCs for second language learning. Within the red cluster, the work of Zeng et al. (2022), on the analysis of collective attention in LMOOCs, with higher link intensity (total link strength 7) and 10 citations, stands out. Additionally, in the green cluster, the work of Peng and Jiang (2022), on the impressions and feelings of Chinese LMOOC students through forums, with medium levels of intensity (total link strength 6) and citations (4), should be highlighted.

3.3. Co-citation and co-occurrence analysis

By applying a co-citation analysis, we know the frequency in which different publications are cited together. This allows us to know the main topics within the object of study. If we add the analysis of co-occurrence of keywords, we can know the main descriptors with respect to the content of the sample of publications. Co-occurrence allows us to connect the frequency of occurrence of keywords with a conceptual link between them. These two analyses complement each other to answer the question about the main lines of research related to LMOOCs in China.

Regarding co-citation, it was established as a criterion to have a minimum of 15 citations, with 16 items fulfilling it (Figure 4). Three sets of co-citations were generated, according to the authors who appear co-cited. The levels of co-citation intensity are diverse. Martin (total link strength 286), Li, C. (total link strength 227) Perun (total link strength 196) and Barcena (total link strength 192) are the most outstanding.
Regarding the co-occurrence of keywords, of the 134 documents in the sample analyzed, the authors proposed 382 keywords and the documents were indexed with 678 keywords, reaching a total of 1060. Among them, 49 keywords coexisted more than five times in the analyzed documents (Figure 5). Five co-occurrence clusters were established, as a result of the linkage between descriptors that usually appear together as keywords.

**FIGURE 4.** Co-citation with “authors” as unit of analysis.

**FIGURE 5.** Concurrence of keywords in scientific production.
Among the descriptors with the greatest presence, and in respect of the Booleans used, MOOC and Massive Open Online Course together reach 89 appearances, while “Language” as such has no presence, although there is mention of “natural language processing” (22), “foreign language” (19), “English teaching” (5), “natural languages” (5), “high level languages” (6) and “language learning” (5). Other terms such as “e-learning” (49) should be highlighted, thus underlining the relationship between online learning and MOOCs as a resource for such processes. Other terms, such as “social networking” (8), “learning behavior” (6), “flipped classroom” (6) and “SPOC” (5) offer an insight into tools, contents and methodologies implemented in LMOOCs that can be further explored in future studies.

4. DISCUSSION

Due to the worldwide development of LMOOCs, in China it has become a relevant field of applied linguistics, which is increasingly attracting attention (Chen, 2022). With adapted MOOC pedagogy, better platforms, easier access to technology, and a general increase in digital literacy, students from different backgrounds can enjoy learning through LMOOCs in line with their own needs and learning pace (Li et al., 2022a). In addition, since language teaching and learning differ from other disciplines, it is essential to pay attention to the technical and management aspects of LMOOC courses (Ye, & Luo, 2021).

Going deeper into the research questions, regarding the evolution of scientific production, we can state that although the number of publications has increased since 2014, they have tended to stabilize around 13-15 publications per year, except as an isolated fact over the two years in which publications increased due to COVID-19. In the sample, technological publications (computer science and engineering) predominate, as they are the basis for the design and development of LMOOCs, followed by social sciences. Regarding the type of documents, contributions to conferences (77) and articles (54) make up the bulk of the sample (97.8%). In terms of affiliation, the most prolific institutions registered five publications (Huazhong Normal University and Tsinghua University), followed by Peking University with four. The most relevant publications, according to the number of citations received, are focused on systematic reviews on LMOOCs spanning from 2009 to 2018 (Fang et al., 2022; Sallam et al., 2022), on the analysis of students’ autonomy to perform LMOOCs and the factors that influence (emotional or motivation, for example) their development (Ding, & Shen, 2022), and on how the implementation of the Flipped Classroom as a methodology in an LMOOC affects the development of students’ communicative competence (Wang et al., 2018).

Regarding the most influential publications on LMOOC in China, the publications by Ding and Shen (2022), Fang et al. (2022) and Sallam et al. (2022) should be highlighted, having already appeared in the most relevant publications (by citations), with the only exception the work by Wang et al. (2018). Thus, systematic reviews on LMOOCs and learner autonomy when using these platforms are references for researchers interested in this topic. Along with these, it is worth highlighting the interest in defining what quality criteria an LMOOC should have for its implementation in learning (Luo, & Ye, 2022), the importance of collective attention when taking this type of second language course (Zeng et al., 2022), as well as the study in forums of the emotional component of taking an LMOOC on the part of the students (Peng, & Jiang, 2022). Thus, the references are recent publications (less than two years old) that focus on learning more about LMOOCs and the students who take them.
Regarding the main lines of research related to LMOOCs in China, it is worth mentioning the combination of MOOCs with different mentions of language learning, such as the work of Jin (2020) for “natural language processing”; He (2022), Liang and Pang (2019) and Wang (2019) for “English teaching”; or Pan et al. (2022) for “foreign learning”. In addition, there are several interesting avenues to explore, such as the use of SPOCs as an evolution of MOOCs (Zhang et al., 2021) or the influence of emotions on student behavior when learning with these methods (Li et al., 2022b). Special mention should be made of the studies in which active methodologies are implemented, such as the flipped classroom, to improve MOOCs in second language learning. Studies were found in respect of the perception of teachers when implementing a MOOC for learning a second language through Flipped Classroom (Orsini, & Zou, 2019); the influence of learning with computers and ICT through MOOCs and Flipped Classroom (Deng, 2021); or a comparison between the possibilities of MOOCs and this methodology for learning a foreign language by analyzing variables such as anxiety, student attitudes or motivation (Pan et al., 2022).

5. CONCLUSIONS

The acquisition of a second language is key in an increasingly global and borderless world. In this technological era, there are many resources, materials and programs available to develop this formative learning. In this situation, LMOOCs become a relevant resource due to their universal and massive access (Martín, & Bárcecena, 2014), being a learning opportunity with ever-increasing possibilities (Jitpaisarnwattana et al., 2021).

In the Chinese context, within the focus of this study, these have been positively accepted by all the agents involved in educational processes (Han, 2019), hence the interest in knowing the evolution of scientific production in this field. From the bibliometric analysis we can conclude that publications increased since the emergence of LMOOCs in 2013 until stabilizing, undergoing a temporary growth during the COVID-19 pandemic. Systematic reviews are notably frequent in this field and are amongst the most cited publications. This is related to their short conceptual trajectory, which is useful for other researchers to situate the theoretical basis of their study proposals. The incorporation of active methodologies into LMOOCs has also had an impact, highlighting the Flipped Classroom as a didactic strategy to be implemented, or as a counterpoint to MOOCs for language learning. In both cases, it is a methodology that arouses interest among researchers and that can continue to be analyzed in the future. The other most common component has to do with the students participating in LMOOCs, analyzing aspects such as their motivation, their autonomy during the learning process or their emotional responses when they make use of this resource. In this sense, it would be interesting to develop specific technologies to improve student interaction and offer them more channels and opportunities for participation in view of the few studies that have analyzed this phenomenon. Moreover, based on the results, if we want to gain a more complete view of the pedagogical impact of LMOOCs, there are other focuses that should be addressed, such as the teaching staff and the quality of the platforms.

5.1. Limitations and foresight

The main limitation of a bibliometric study is not incorporating publications from other international data- bases such as Web of Science, Eric or Scielo. The main factor is related to the duplication of publications,
due to the fact that journals tend to be indexed simultaneously in several databases. In addition, we acknowledge that the criteria established by Scopus for indexing in its database are of sufficient quality for the sample found to be representative of the phenomenon under study.

Among the future lines of research, if we modify the inclusion criteria, it would be interesting to analyze the impact of LMOOCs on scientific production worldwide, which would provide us with information on the most prolific countries and institutions, together with the main research groups and lines of work on this topic. As a result of this and of the detailed analysis of the most relevant research worldwide (systematic review as a working methodology), the strengths and weaknesses of LMOOCs could be outlined, and constructive suggestions could be made for the future design of technology-assisted language learning.

6. FUNDING

This research has been funded in part by the CQI-Valencia Innovation Experience Chair of the Universitat Politècnica de València, grant number CON2022-111402-R-CQI, Red CQI Promotion. The views expressed in this document are not necessarily those of the organization.

7. REFERENCES


Language MOOCs: Providing Learning, Transcending Boundaries (pp. 1–15). De Gruyter Open Poland. [https://doi.org/10.2478/9783110420067.1]


