International Review of Research in Open and Distributed Learning

Investigation of Emerging Trends in the E-Learning Field Using Latent Dirichlet Allocation

Fatih Gurcan, Ozcan Ozyurt and Nergiz Ercil Cagitay

Cite this article

Gurcan, F., Ozyurt, O. & Cagitay, N. (2021). Investigation of Emerging Trends in the E-Learning Field Using Latent Dirichlet Allocation. *International Review of Research in Open and Distributed Learning*, *22*(2), 1–18. https://doi.org/10.19173/irrodl.v22i2.5358

Article abstract

E-learning studies are becoming very important today as they provide alternatives and support to all types of teaching and learning programs. The effect of the COVID-19 pandemic on educational systems has further increased the significance of e-learning. Accordingly, gaining a full understanding of the general topics and trends in e-learning studies is critical for a deeper comprehension of the field. There are many studies that provide such a picture of the e-learning field, but the limitation is that they do not examine the field as a whole. This study aimed to investigate the emerging trends in the e-learning field by implementing a topic modeling analysis based on latent Dirichlet allocation (LDA) on 41,925 peer-reviewed journal articles published between 2000 and 2019. The analysis revealed 16 topics reflecting emerging trends and developments in the e-learning field. Among these, the topics "MOOC," "learning assessment," and "e-learning systems" were found to be key topics in the field, with a consistently high volume. In addition, the topics of "learning algorithms," "learning factors," and "adaptive learning" were observed to have the highest overall acceleration, with the first two identified as having a higher acceleration in recent years. Going by these results, it is concluded that the next decade of e-learning studies will focus on learning factors and algorithms, which will possibly create a baseline for more individualized and adaptive mobile platforms. In other words, after a certain maturity level is reached by better understanding the learning process through these identified learning factors and algorithms, the next generation of e-learning systems will be built on individualized and adaptive learning environments. These insights could be useful for e-learning communities to improve their research efforts and their applications in the field accordingly.

Investigation of Emerging Trends in the E-Learning Field Using Latent Dirichlet Allocation

Fatih Gurcan[1]*, Ozcan Ozyurt[2], and Nergiz Ercil Cagiltay[3]
[1]*Distance Education Application and Research Centre, Karadeniz Technical University, Trabzon, Turkey;* [2]*Karadeniz Technical University, OF Technology Faculty, Software Engineering Department, Trabzon, Turkey;* [3]*Software Engineering Department, Atilim University, Ankara, Turkey*
*Corresponding Author

Abstract

E-learning studies are becoming very important today as they provide alternatives and support to all types of teaching and learning programs. The effect of the COVID-19 pandemic on educational systems has further increased the significance of e-learning. Accordingly, gaining a full understanding of the general topics and trends in e-learning studies is critical for a deeper comprehension of the field. There are many studies that provide such a picture of the e-learning field, but the limitation is that they do not examine the field as a whole. This study aimed to investigate the emerging trends in the e-learning field by implementing a topic modeling analysis based on latent Dirichlet allocation (LDA) on 41,925 peer-reviewed journal articles published between 2000 and 2019. The analysis revealed 16 topics reflecting emerging trends and developments in the e-learning field. Among these, the topics "MOOC," "learning assessment," and "e-learning systems" were found to be key topics in the field, with a consistently high volume. In addition, the topics of "learning algorithms," "learning factors," and "adaptive learning" were observed to have the highest overall acceleration, with the first two identified as having a higher acceleration in recent years. Going by these results, it is concluded that the next decade of e-learning studies will focus on learning factors and algorithms, which will possibly create a baseline for more individualized and adaptive mobile platforms. In other words, after a certain maturity level is reached by better understanding the learning process through these identified learning factors and algorithms, the next generation of e-learning systems will be built on individualized and adaptive learning environments. These insights could be useful for e-learning communities to improve their research efforts and their applications in the field accordingly.

Keywords: e-learning, text mining, topic modeling, trends, developmental stages

Introduction

Today, e-learning has become a very important topic, with applications in every field, as supportive training, lifelong learning modalities, and support tools, for all types of educational systems. Due to the effects of the COVID-19 pandemic on teaching and learning environments, research into e-learning studies has become even more critical. A recent study by Chavarría-Bolaños et al. (2020) reported the importance of e-learning in dental education, for example. E-learning studies can be considered as multidisciplinary, as several fields contribute to it from different perspectives. The roots of e-learning studies go back to the late 1950s, and therefore, there is a large amount of available literature detailing improvements and achievements in this field over the decades. Furthermore, as highlighted by some researchers, since 2000, the number of studies conducted on e-learning has significantly increased (González, 2010) and will likely accelerate in the current pandemic situation. By analyzing these studies, one can get a general overview of e-learning studies that can help us understand how the field is evolving and where it is going. Such studies are very critical in guiding future research and developments related to all kind of e-learning studies. In the literature, there have been several attempts to analyze earlier studies and provide a general overview of the field. As defined by Rowley and Slack (2004), systematic reviews aim to facilitate the definition, evaluation, and interpretation of studies in a specific field by examining the concepts, applications, and theories pertaining to it. These studies systematically review the literature to answer research questions to better understand and examine the key concepts in the field. Some of the previous studies on e-learning were conducted to provide insights into a specific area of e-learning, such as the Semantic Web for distance learning (Bashir & Warraich, 2020), virtual education (Fermín-González, 2019), educational data mining (Rodrigues et al., 2018), mobile learning in higher education (Krull & Duart, 2017), and machine-learning-based recommendation systems for e-learning (Khanal et al., 2020). Another group of studies were conducted on the implementation of e-learning in specific fields, such as e-learning for training work corporations (Kaizer et al., 2020), e-learning in undergraduate dentistry education (Zitzmann et al., 2020), implications of e-learning for universities (Kibuku et al., 2020), and e-learning for mathematics teaching (Klingenberg et al., 2020). There are only a limited number of systematic review studies addressing e-learning studies in general. Among these, a systematic review conducted on 99 e-learning articles published between 2010 and 2018 reported four main themes in the field: educational systems, learning issues, student behaviors, and online learning tools (Rodrigues et al., 2019). Valverde-Berrocoso et al. (2020) also conducted a systematic review analyzing 248 articles published between 2009 and 2018 and discovered the following: that online students, online teachers, and curriculum-interactive learning environments were the three main nodes of e-learning; that MOOCs were the most researched e-learning modality; that the community of inquiry and the technological acceptance model were the most used theories in the analyzed studies; and, finally, that case studies were the most frequently used methodology. As these systematic reviews require a lot of researcher effort, they are usually conducted with a limited number of articles.

Another group of studies attempting to provide a bigger picture of e-learning studies were undertaken as bibliometric analyses in scientific and research fields to examine the properties and recorded information based on a number of indicators (Abramo et al., 2009; Patra et al., 2006). As these studies considered certain indicators as the basis for analysis, they were conducted on larger data sets. For instance, Hung (2012) examined 689 articles published between 2000 and 2008 through a bibliometric analysis. Similarly, Asadzandi et al. (2017) descriptively analyzed 23,805 e-learning studies through the categories provided by the Scopus database, such as date of publication, type of documents, language of the documents, source of

articles, subject areas, authors, and their affiliations, concluding that there was a steady growth in the number of articles on e-learning studies, which was parallel with its development. Similarly, Tibaná-Herrera et al. (2018a) categorized e-learning as an emerging discipline consisting of 64 descriptors and 219 journals and congresses indexed by Scopus between 2012 and 2014. Another bibliometric analysis was conducted by Tibaná-Herrera et al. (2018b) on 39,244 documents published between 2003 and 2016 that were indexed by Scopus and SCImago Institutional Rankings. They reported the following: the majority of these studies were published by authors from the United States; the University of Hong Kong was the most productive institution; and the National Taiwan University of Science and Technology had the greatest collaboration. Thus, bibliometric analyses were conducted on larger data sets and possibly provide a bigger picture of the field; however, as the analysis was based on a number of indicators, these bibliometric analyses missed out the details in the content of the published studies, which limits their contributions to the field.

All of these earlier studies are very valuable in providing a general perspective of the field of e-learning, despite limitations such as the limited number of articles, the narrow scope of the field, or limitations in the analysis methods (Çakiroğlu et al., 2019). As the number of articles in the field of e-learning is significantly increasing, it is becoming more difficult to conduct a manual analysis (Yang et al., 2016). Different methods are used for in-depth analysis of superficial description. In this context, various analyses can be performed using text/data mining methods with a large number of article sets. Today, different types of text analysis of a high volume of documents, such as word frequency analysis, text classification approaches, topic modeling analysis, and n-gram analysis, are being used extensively to gain a deeper understanding of specific domains and fields (Gurcan, 2019; Gurcan et al., 2021). For instance, in the field of distance education, Gurcan and Cagiltay (2020) recently conducted a text-mining-based review by analyzing 27,735 peer-reviewed journal articles published between 2008 and 2018 using n-grams, and they reported 10 main themes of the field. However, they applied a manual classification on the topics identified (Gurcan & Cagiltay, 2020). Recently, with improvements in machine learning and data mining techniques, significant developments have occurred in the areas of automatic topic determination, semantic information extraction from texts, and automatic analysis of very large data sets using text mining methodologies (Gürcan, 2009; Gurcan, 2018). These techniques open a wider window into understanding studies in the field and offer objective analysis methods. Accordingly, the study discussed in this article aimed to provide a wider perspective by analyzing 41,925 e-learning journal articles and reviews published between 2000 and 2019 using the latent Dirichlet allocation (LDA) algorithm (Blei et al., 2003). The methodology of the study was designed to investigate the following research questions (RQ):

RQ1. What have been the bibliometric characteristics of e-learning research during the period between 2000 and 2019?

RQ2. What have been the emerging topics in the e-learning field in the period between 2000 and 2019?

RQ3. How have the topics of interest in e-learning studies changed from 2000 to 2019?

RQ4. What are the future trends in the e-learning field?

Methods

The literature available on e-learning is very comprehensive. Since journal articles are subjected to a peer review process, this study considered only peer-reviewed journal articles. More specifically, only e-learning-oriented journal articles published in English in the last 20 years (between 2000 and 2019) were included in this study. Since e-learning is an interdisciplinary field covering a wide spectrum of topics, an iterative strategy was followed to determine the search string for the study. Namely, first, a wide literature review was carried out in order to determine the synonym equivalents of e-learning expression in the literature. Then, the opinions of field experts were obtained regarding the extracted terms. The final keywords were determined from the results of the examination by five field experts and the evaluation of the researchers. The search query that met the search string and other criteria determined as a result of these processes was created as follows:

> TITLE-ABS-KEY (("online learning" OR "e-learning" OR "distance learning" OR "mobile learning" OR "web-based learning" OR "online training" OR "e-training" OR "distance training" OR "mobile training" OR "web-based training" OR "online education" OR "e-education" OR "distance education" OR "mobile education" OR "web-based education" OR "online teaching" OR "e-teaching" OR "distance teaching" OR "mobile teaching" OR "web-based teaching" OR "MOOC" OR "online open course")) AND (PUBYEAR < 2020) AND (PUBYEAR > 1999) AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "re")) AND (LIMIT-TO (LANGUAGE , "English"))

The Scopus database was used to obtain articles suitable for the scope of the study since it covers more than 5000 publishers worldwide—including Elsevier, Emerald, IEEE, Sage, Springer, Taylor & Francis, and Wiley Blackwell—and this number is increasing daily (Gurcan et al., 2021; Mongeon & Paul-Hus, 2016). The query given above was run on April 5, 2020, to access the relevant articles from the Scopus database. The search brought up a total of 41,925 articles (2619 review articles and 39,306 research articles). The title, abstract, and author keyword information of these articles were added to the data set.

To prepare the e-learning corpus for probabilistic topic modeling, preprocessing tasks such as tokenization; removing meaningless words, symbols, and stop words; and stemming were implemented (Gurcan et al., 2021). Then, an e-learning document term matrix was created, in which each row represented an article and each column represented a unique word in the e-learning corpus. Afterward, LDA, a probabilistic topic modeling approach (Blei et al., 2003), was used for creating and fitting a topic model to the e-learning corpus and analyzing this corpus.

LDA is a generative approach used to discover hidden semantic patterns in a large, relatively unstructured document corpus (Blei, 2012). Text documents contain hidden semantic patterns called "topics," and each of these topics is defined by a probability distribution over a fixed set of words (Blei et al., 2003). Since LDA is an unsupervised method for topic modeling, it does not require any training set, tags, or metadata for learning, so large numbers of textual documents can be analyzed in a short time. The LDA model is frequently used in content analysis based on topic modeling (Blei et al., 2003; Blei & Lafferty, 2007). For these reasons, the LDA model was preferred over others and employed for topic modeling analysis of the e-learning corpus in this study.

This analysis revealed 16 topics at an optimal level. The top 20 words with the highest probability were identified for each topic and assigned to these topics. A suitable topic name was defined for each topic taking into account the first five words in the topics. Furthermore, the volumetric percentage rates and the temporal trends of the topics that modeled the entire e-learning corpus were revealed by calculating the distribution of topics per document and the word distributions per topic (Gurcan et al., 2021; Gurcan & Cagiltay, 2020).

Results

The results of the study are first presented descriptively by considering the number of yearly publications, the top subject areas and journals, and the top countries of the authors. Additionally, the top keywords found in these articles are also mentioned descriptively. Further, a detailed topic modeling analysis is presented to provide an overall picture of e-learning studies.

Descriptive Analysis

In order to describe the bibliometric characteristics of the e-learning field between 2000 and 2019 (RQ1), the descriptive analysis of the corpus is given below. The total number of articles published between 2000 and 2019 and their yearly distribution are given in Table 1, showing a total of 41,925 articles analyzed in the study. It should be noted that although there was a slight decrease in the number of articles in 2002 and 2010 compared to the other years, there was an overall linear increase in the number of publications each year.

Table 1

Yearly Distribution of the Articles

Year	n	%
2000	681	1.62
2001	861	2.05
2002	788	1.88
2003	993	2.37
2004	1133	2.70
2005	1179	2.81
2006	1325	3.16
2007	1508	3.60
2008	1632	3.89
2009	1962	4.68
2010	1927	4.60
2011	2244	5.35
2012	2330	5.56
2013	2494	5.95
2014	2847	6.79
2015	3017	7.20
2016	3224	7.69
2017	3357	8.01

2018	3823	9.12
2019	4600	10.97
Total	41,925	100

Figure 1 shows the top 10 subject areas addressed by the highest number of articles. The majority of the articles were published in the field of social sciences, including educational sciences ($n = 23{,}150$). As some studies were carried out in more than one discipline, they were classified under each of these subject areas by Scopus.

Figure 1

Top 10 Subject Areas With the Most Published Articles
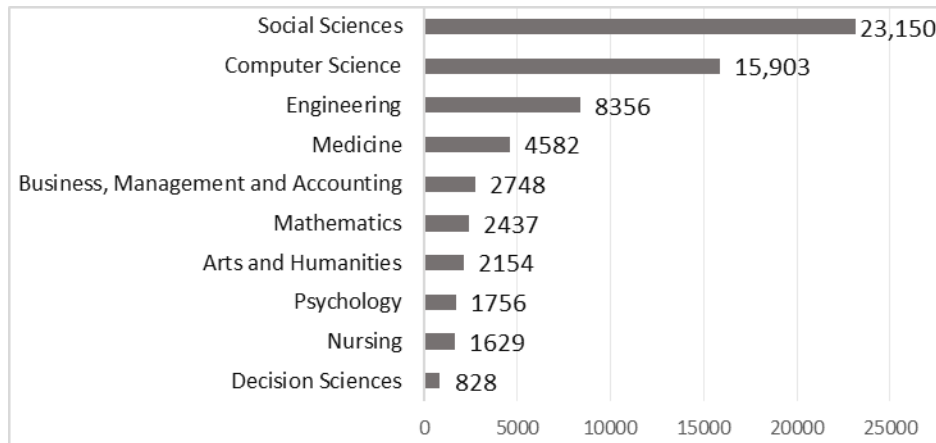
Figure 2 shows the top 10 journals with the highest number of published articles. The *Computers and Education* journal published the highest number of articles ($n = 975$), followed by the *International Review of Research in Open and Distance Education* journal ($n = 723$) and the *Turkish Online Journal of Distance Education* journal ($n = 688$).

Figure 2

Top 10 Journals With the Most Published Articles
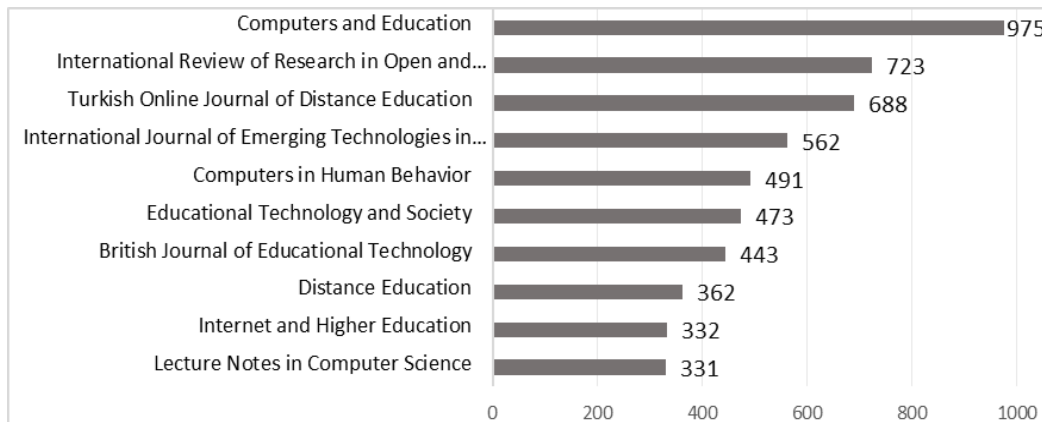
Figure 3 reveals that the highest number of articles originated from the United States of America ($n = 12{,}024$; f = 28.7%), followed by the United Kingdom ($n = 3950$) and China ($n = 3223$).

Figure 3

Top 10 Countries With the Most Published Articles
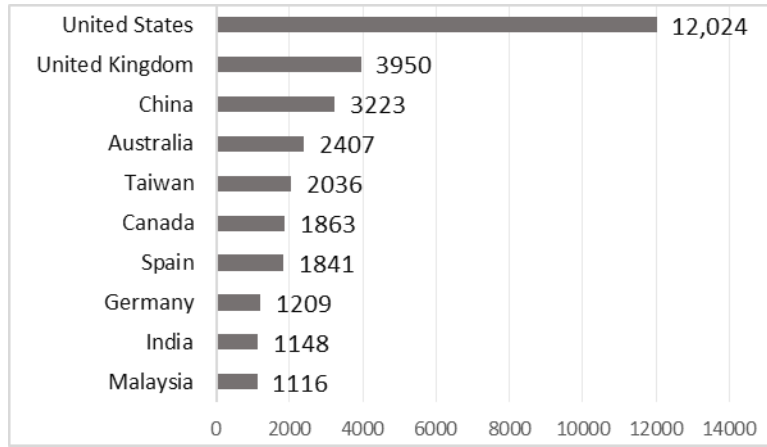
The top 20 keywords of the analyzed studies are listed in Table 2, with the top five keywords being "e-learning" (30.68%), "human" (27.35%), "education" (16.42%), "teaching" (12.88%), and "student" (12.01%).

Table 2

Top 20 Keywords Addressed by E-Learning Articles

Keyword	n	%
E-learning	12,861	30.68
Human	11,466	27.35
Education	6885	16.42
Teaching	5402	12.88
Student	5034	12.01
Distance education	4591	10.95
Online learning	4144	9.88
Internet	3935	9.39
Learning	3145	7.50
Learning systems	3111	7.42
Female	2573	6.14
Male	2423	5.78
Computer aided instruction	2196	5.24
Distance learning	2192	5.23
Adult	2098	5.00
Medical education	2010	4.79
Mobile learning	1692	4.04
Higher education	1560	3.72
Online systems	1546	3.69
Curriculum	1487	3.55

Topic Modeling Analysis

In order to reveal the emerging topics in the e-learning field (RQ2), the results of the topic modeling analysis achieved by the LDA are given in this section. Using a LDA-based topic modeling procedure, 16 topics were discovered (see Table 3). The rate (%) of each topic was calculated by their volume, referring to the number of articles published on each topic.

Table 3

Discovered Topics and Keywords of the Articles

Topic name	Keywords	Rate %
MOOC	learn*, educ*, onlin*, design*, mooc*, develop*, practic*, approach*, teach*, cours*, learner*, technologi*, experi*, environ*, theori*, model*, support*, activ*, framework*, context*	10.13
Learning assessment	student*, learn*, cours*, onlin*, assess*, effect*, result*, teach*, instruct*, perform*, blend*, class*, feedback*, evalu*, tradit*, classroom*, compar*, method*, lectur*, test*	9.86
Distance education	educ*, distanc*, student*, univers*, program*, onlin*, cours*, learn*, faculti*, technologi*, develop*, institut*, teach*, academ*, support*, offer*, graduat*, experi*, access*, colleg*	9.68
E-learning systems	elearn*, system*, learn*, educ*, develop*, manag*, technologi*, evalu*, model*, process*, design*, tool*, qualiti*, inform*, implement*, applic*, support*, platform*, univers*, environ*	9.05
Learning algorithms	learn*, onlin*, algorithm*, model*, network*, control*, method*, data*, system*, perform*, neural*, predict*, adapt*, result*, optim*, train*, featur*, machin*, track*, dynam*	9.02
Educational management	train*, develop*, manag*, countri*, educ*, project*, polici*, cultur*, global*, busi*, knowledg*, chang*, organ*, inform*, market*, commun*, industri*, sustain*, intern*, employe*	6.34
Adaptive learning	learn*, system*, learner*, adapt*, data*, model*, user*, knowledg*, person*, intellig*, approach*, elearn*, recommend*, style*, object*, content*, environ*, mine*, semant*, result*	6.00
Medical education	health*, medic*, educ*, nurs*, care*, train*, clinic*, patient*, practic*, knowledg*, method*, profession*, program*, particip*, evalu*, develop*, improv*, skill*, modul*, assess*	5.92
Social learning	learn*, commun*, onlin*, social*, collabor*, interact*, discuss*, student*, environ*, particip*, network*, presenc*, support*, activ*, media*, share*, virtual*, knowledg*, forum*, asynchron*	5.91
Learning factors	factor*, learn*, student*, model*, perceiv*, motiv*, influenc*, result*, satisfact*, effect*, accept*, attitud*, learner*, analysi*, data*, intent*, affect*, technologi*, signific*, percept*	5.90
Virtual systems	virtual*, system*, engin*, laboratori*, comput*, simul*, environ*, remot*, experi*, design*, interact*, control*, develop*, applic*, cloud*, train*, realiti*, network*, educ*, technologi*	5.38
Information resources	inform*, video*, librari*, resourc*, web*, servic*, digit*, access*, internet*, content*, literaci*, materi*, user*, site*, websit*, search*, librarian*, educ*, lectur*, multimedia*	4.34
Training	train*, intervent*, effect*, control*, particip*, result*, improv*, trial*, children*, health*, program*, test*, assess*, patient*, increas*, measur*, outcom*, compar*, behavior*, prevent*	3.80

Mobile learning	mobil*, learn*, devic*, game*, technologi*, educ*, student*, mlearn*, applic*, app*, develop*, phone*, digit*, activ*, support*, design*, environ*, smart*, comput*, result*	3.74
Language teaching	languag*, teach*, learn*, english*, multimedia*, learner*, write*, effect*, student*, technologi*, read*, skill*, develop*, improv*, comput*, platform*, foreign*, design*, chines*, applic*	3.16
Teacher education	teacher*, school*, educ*, mathemat*, teach*, ict*, profession*, develop*, train*, secondari*, classroom*, pre-servic*, technologi*, compet*, scienc*, primari*, music*, elementari*, mentor*, digit*	1.76

The top 20 keywords classified under each topic are also given by considering their volume rates. Table 3 shows that the most intensively studied topic by researchers was "MOOC" (10.13%), while the least read topic was "teacher education" (1.76%). Figure 4 shows the volume of the topics among all the articles considered in this study. Accordingly, the topics can be classified as high-volume topics having a ratio higher than 9.0%, medium-volume topics having a ratio higher than 5.4% and less than 9.0%, and low-volume topics having a ratio less than 5.4%. The topics having the highest ratio were "MOOC" (10.13%), "learning assessment" (9.86%), "distance education" (9.68%), "e-learning systems" (9.05%), and "learning algorithms" (9.02%), while those with lower ratios were "teacher education" (1.76%), "language teaching" (3.16%), "mobile learning" (3.74%), "training" (3.80%), and "information resources" (4.34%). According to these ratio differences, the discovered topics could be classified under three groups. Changes in the volume ratios were taken into account while classifying the discovered topics. There were sharp decreases and clusters in volume ratios below 9 and below 5. These groups were labeled by the researchers as high-volume (n = 5), medium-volume (n = 6), and low-volume (n = 5) topics.

Figure 4

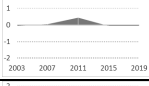*Percentage Rates of Articles From 2000 to 2019 for Each Topic*
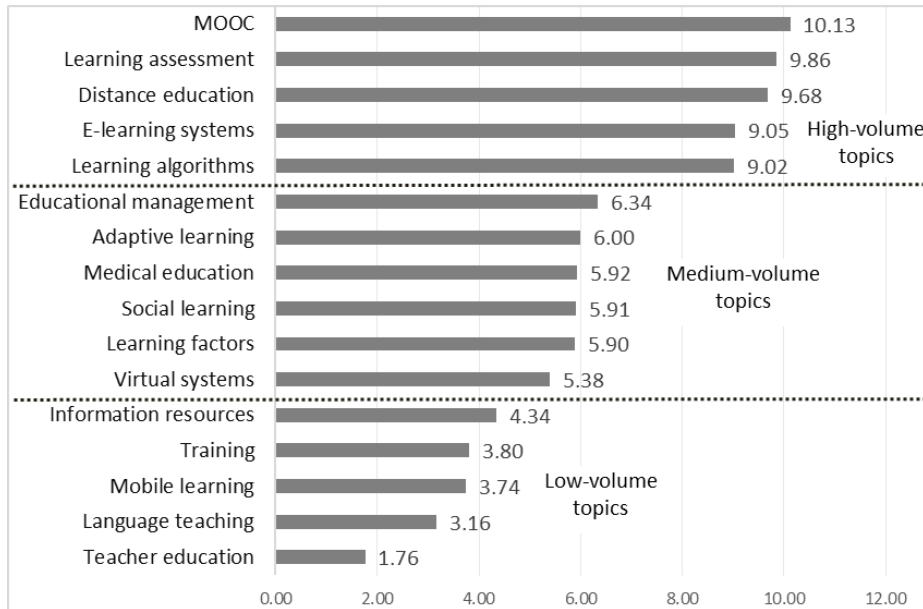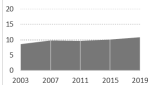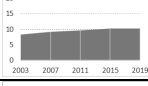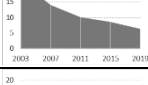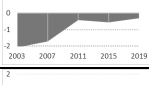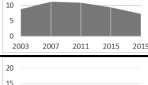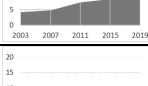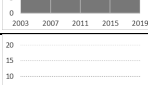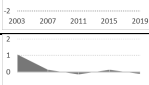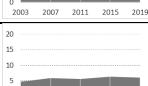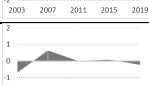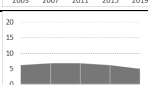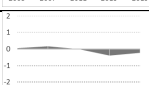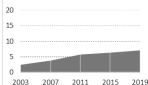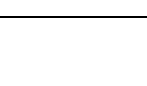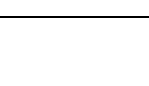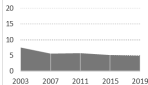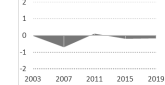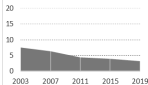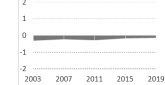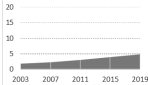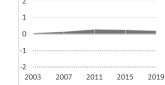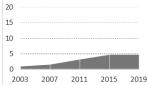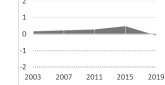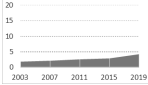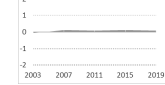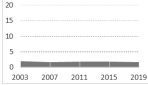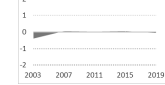
To better understand the temporal trends of e-learning topics between 2000 and 2019 (RQ3), the developmental stages of these topics were analyzed in four-year periods as shown in Table 4, with the

average number of articles published under each topic (n) for each time period being evaluated. Their percentages according to the total number of articles published each year were calculated, and their average value for each period (%) is also given. Their accelerations were calculated by subtracting the average percentage of articles from that of the previous years. The average acceleration values (A) for each period were also calculated and are presented in Table 4. Finally, the trends of the articles for each topic are presented graphically, considering their volume according to the percentages of the number of articles (%) and the acceleration graph through the calculated acceleration values (A). Table 4 shows that among the top volume topics, "MOOC" and "learning assessment" showed more steady behavior; however, for some topics, such as "distance education," there was a decrease and for other topics, such as "learning algorithms," there was an increase in the percentages of the number of periodical articles. Similarly, even though "teacher education" had the lowest volume, it had a steady acceleration resulting in a similar number of articles compared to the other topics.

Table 4

Volume and Acceleration of Articles for Each Discovered Topic in Four-Year Periods

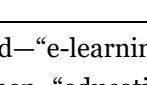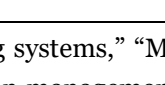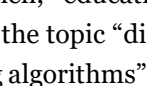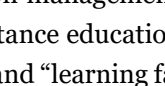| Topic name		2000–2003	2004–2007	2008–2011	2012–2015	2016–2019	AVG	Volume graph (%)	Acceleration graph
MOOC	n	285	503	750	1092	1619	849.74		
	%	8.62	9.75	9.64	10.17	10.83	9.80		
	A	−0.31	0.50	0.00	0.19	−0.14	0.04		
Learning assessment	n	278	474	752	1093	1536	826.47		
	%	8.37	9.25	9.62	10.22	10.26	9.54		
	A	−0.05	0.05	0.44	−0.05	−0.05	0.07		
Distance education	n	708	707	779	908	955	811.30		
	%	21.59	13.96	10.07	8.58	6.40	12.12		
	A	−2.07	−1.70	−0.42	−0.56	−0.31	−1.01		
E-learning systems	n	298	579	842	990	1085	758.68		
	%	8.78	11.19	10.86	9.34	7.28	9.49		
	A	1.33	0.12	−0.38	−0.51	−0.33	0.05		
Learning algorithms	n	142	253	577	919	1893	756.74		
	%	4.23	4.86	7.44	8.50	12.36	7.48		
	A	0.14	0.32	0.34	0.83	1.21	0.57		
Educational management	n	319	438	521	613	767	531.80		
	%	9.47	8.49	6.75	5.77	5.12	7.12		
	A	0.96	−0.36	−0.68	−0.34	0.04	−0.08		
Adaptive learning	n	127	332	529	656	873	503.20		
	%	3.66	6.47	6.87	6.12	5.80	5.78		
	A	1.06	0.15	−0.14	0.15	−0.11	0.22		
Medical education	n	153	302	440	676	911	496.55		
	%	4.72	5.86	5.62	6.30	6.10	5.72		
	A	−0.69	0.64	0.01	0.09	−0.21	−0.04		
Social learning	n	204	348	524	656	746	495.68		
	%	6.13	6.75	6.74	6.19	5.04	6.17		
	A	0.05	0.16	−0.01	−0.40	−0.22	−0.08		
Learning factors	n	83	199	444	680	1066	494.38		
	%	2.47	3.81	5.68	6.36	7.05	5.07		
	A	0.15	0.53	0.37	0.07	0.30	0.28		
	n	252	288	439	550	728	451.48		

Virtual systems	%	7.59	5.60	5.67	5.16	4.90	5.79		
	A	−0.04	−0.69	0.12	−0.18	−0.18	−0.20		
Information resources	n	249	323	344	418	485	363.96		
	%	7.55	6.34	4.45	3.94	3.25	5.11		
	A	−0.31	−0.23	−0.29	−0.16	−0.13	−0.22		
Training	n	63	122	236	432	742	319.03		
	%	1.89	2.35	3.01	4.01	4.90	3.23		
	A	0.06	0.13	0.27	0.24	0.21	0.18		
Mobile learning	n	33	81	251	504	699	313.73		
	%	0.97	1.54	3.24	4.67	4.65	3.01		
	A	0.16	0.22	0.27	0.47	−0.09	0.21		
Language teaching	n	63	111	199	312	640	264.79		
	%	1.90	2.15	2.54	2.90	4.30	2.76		
	A	−0.05	0.12	0.08	0.11	0.07	0.07		
Teacher education	n	66	84	139	190	259	147.46		
	%	2.06	1.64	1.79	1.78	1.74	1.80		
	A	−0.38	0.06	0.04	0.06	−0.05	−0.06		

Note. A = acceleration; AVG = average.

When analyzed, the top topics considering their volumes in each period—"e-learning systems," "MOOC," and "learning assessment"—were in the top five from 2000 to 2020; then, "education management" was one of the top five topics between 2000 and 2008 (Figure 5). Similarly, the topic "distance education" was one of the top five topics starting from 2000 to 2016. The topics "learning algorithms" and "learning factors" appeared on the list starting from 2008 and 2016, respectively.

Figure 5

The Top Five Topics From 2000 to 2020
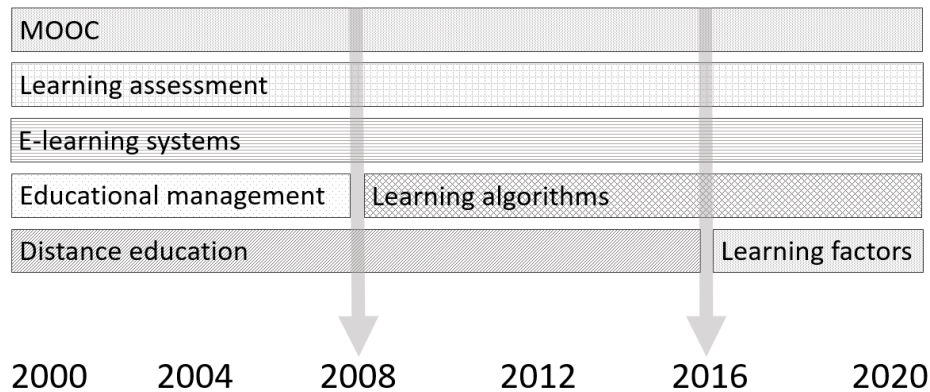
In order to reveal insights about future trends in e-learning (RQ4), the acceleration of the discovered topics in terms of their average acceleration for all years is given in Figure 6. "Learning algorithms" had the highest acceleration value (0.57) and "distance education" had a significantly lower acceleration value (−1.01) compared to the other topics.

Figure 6

Acceleration of Topics From 2000 to 2020
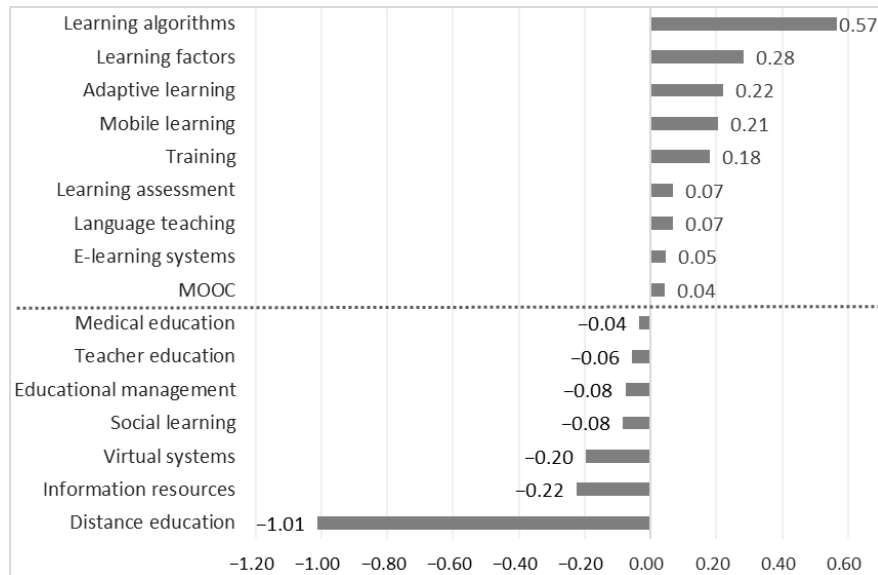
The recent trends of the topics and their acceleration values during the last period (2016–2019) are given in Figure 7. "Learning algorithms" had a significantly higher acceleration (1.21), and during the same period, the acceleration of the topics "e-learning systems" (−0.33) and "distance education" (−0.31) was the lowest.

Figure 7

Acceleration of Topics From 2016 to 2020
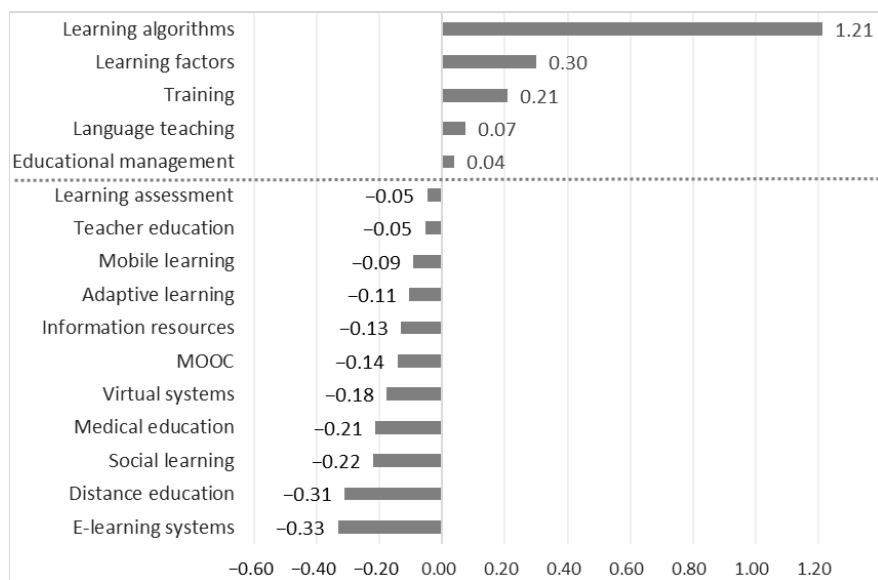
Discussion

In this study, the main trends of e-learning during the last 20 years (between 2000 and 2019) were determined by analyzing articles published in the field using a topic modeling analysis, and 16 main topics were discovered through the LDA-based analysis. The number of articles in this field showed a linear increase over the years (see Table 1), a result parallel with earlier work reporting that studies in the field of e-learning have started to increase and become widespread especially since the early 2000s (Tibaná-Herrera et al., 2018a). The results revealed that the top five subject areas were social sciences, computer science, engineering, medicine and business, and management and accounting. Considering that educational science is also under social sciences, our results were aligned with those of Tibaná-Herrera et al. (2018a), indicating educational science as being the major subject area for e-learning studies. Additionally, by highlighting "medical education" as one of the discovered topics (see Table 3), the results of the current study support earlier work suggesting that in recent years, e-learning studies in the field of medicine are in first place (Barteit et al., 2020). According to the results, in the e-learning corpus, the majority of the articles (975 of them) were published in the *Computers & Education* journal, which indicates that this journal creates a larger space for e-learning studies (see Figure 2). An examination of the origins of the articles showed that the United States was in the lead (see Figure 3; 12,024 articles), which supports the findings of Tibaná-Herrera et al. (2018b). In addition to these contributions, the results of the current study offer insights into e-learning studies, which are summarized under three main headings as follows:

Emergence of New Topics

Table 4 reveals that during the early years (2000–2003) of the publication of e-learning studies, "distance education" (21.59%) had the highest volume ratio and can be considered as the main and oldest topic of e-learning studies. In contrast, during this period, "mobile learning" (0.97%) and "training" (1.89%) had a lower volume ratio in terms of the percentage of articles; thus, they can be classified as having been very young and newly emerging topics in those years. When the acceleration values of these topics were analyzed, as seen in Figure 6, "distance education" had the lowest acceleration value (−1.01), an indicator that the emergence of these younger topics, such as "mobile learning" and "training," decreases the volume percentages of the older topics like "distance education."

Major Topics

The results of this study indicate that "learning algorithms," "learning factors," and "adaptive learning" were the major topics having the highest overall acceleration values (0.57, 0.28, and 0.22, respectively; Figure 6). Additionally, Table 4 shows that the topic "MOOC" had the highest average volume (n = 849.74). These results seem to confirm the expectation of Graf et al. (2010) that MOOCs would occupy an important place in the future. In addition, Chiappe and Lee (2017) supported the view that MOOCs had an important place in e-learning, which is also consistent with the findings of Valverde-Berrocoso et al. (2020) that reported MOOCs as being the most researched e-learning modality.

Future of the Field

The analysis of the accelerations of the topics revealed that after 2008, "learning algorithms" and "learning factors" were also becoming dominating topics with higher overall (0.57 and 0.28, respectively; Figure 6) and recent (1.21 and 0.30, respectively; Figure 7) acceleration values. As in the current stage of e-learning

systems a large amount of data is being collected from e-learning activities, studies on "learning algorithms" and "learning factors" will offer an understanding of the learning process, which will also create a baseline for its adaptation and individualization. As it is not easy to thoroughly create adaptive e-learning systems without developing appropriate learning algorithms and without a deeper understanding of the learning factors, the acceleration of the topic "adaptive learning" has recently dropped from an overall acceleration value of 0.22 (Figure 6) to −0.11 (Figure 7). However, after developments in topics such as "learning algorithms" and "learning factors," the acceleration of "adaptive learning" can be expected to show an increase in the following decades, with a similar trend for "mobile learning."

Conclusion

In this study, 16 main topics of e-learning studies were identified, and the results of the study are important in terms of determining the trends in the field of e-learning. Based on the results of this study, it can be concluded that "learning algorithms," "learning factors," "training," "language teaching," and "educational management" have been the highly accelerating topics during the last four years, and in the near future, they are expected to have an even greater impact on the field and create a baseline for more individualized and adaptive mobile platforms. Accordingly, it can be concluded that although the field is encompassing more adaptive e-learning systems, the developments for supporting adaptive e-learning platforms are not yet sufficiently mature, and during the next few years, the dominating topics will be those five topics. However, after these five topics reach a level of maturity, "adaptive learning" and "mobile" can be expected to have higher acceleration. The results of the current study can offer support to researchers working in this field, as well as to decision-makers and practitioners. In future studies, similar analyses can be conducted to determine the changes in this field and perform comparative studies. Furthermore, the results obtained from this work can lead to more comprehensive studies on sub-topics based on both high-volume and fast-accelerating issues.

In this study, LDA-based topic modeling technique was implemented on 41,925 peer-reviewed journal articles. Even though this technique provides an opportunity to analyze large data sets, currently, it is not possible to conduct deeper analyses like systematic reviews through LDA. In the future, with improvements in topic modeling algorithms, deeper analysis of large data sets can also be performed, which could be expected to provide very important insights for the researchers in this field.

References

Abramo, G., D'Angelo, C. A., & Caprasecca, A. (2009). Allocative efficiency in public research funding: Can bibliometrics help? *Research Policy, 38*(1), 206–215. https://doi.org/10.1016/j.respol.2008.11.001

Asadzandi, S., Rakhshani, T., & Mohammadi, A. (2017). Content analysis study of e-learning literature based on scopus record through 2013: With a focus on the place of Iran's productions. *International Journal on E-Learning: Corporate, Government, Healthcare, and Higher Education, 16*(3), 213–229. https://eric.ed.gov/?id=EJ1140955

Barteit, S., Guzek, D., Jahn, A., Bärnighausen, T., Jorge, M. M., & Neuhann, F. (2020). Evaluation of e-learning for medical education in low- and middle-income countries: A systematic review. *Computers and Education, 145.* https://doi.org/10.1016/j.compedu.2019.103726

Bashir, F., & Warraich, N. F. (2020). Systematic literature review of Semantic Web for distance learning. *Interactive Learning Environments.* https://doi.org/10.1080/10494820.2020.1799023

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM, 55*(4), 77–84. https://doi.org/10.1145/2133806.2133826

Blei, D. M., & Lafferty, J. D. (2007). Correction: A correlated topic model of *Science. The Annals of Applied Statistics, 1*(2), 634. https://doi.org/10.1214/07-aoas136

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research, 3*(4/5), 993–1022. https://dl.acm.org/doi/10.5555/944919.944937

Çakiroğlu, Ü., Kokoç, M., Gökoğlu, S., Öztürk, M., & Erdoğdu, F. (2019). An analysis of the journey of open and distance education: Major concepts and cutoff points in research trends. *International Review of Research in Open and Distance Learning, 20*(1), 2–20. https://doi.org/10.19173/irrodl.v20i1.3743

Chavarría-Bolaños, D., Gómez-Fernández, A., Dittel-Jiménez, C., & Montero-Aguilar, M. (2020). E-learning in dental schools in the times of COVID-19: A review and analysis of an educational resource in times of the COVID-19 pandemic. *Odovtos – International Journal of Dental Sciences, 22*(3), 69–86. https://doi.org/10.15517/ijds.2020.41813

Chiappe, A., & Lee, L. L. (2017). Open teaching: A new way on e-learning? *Electronic Journal of E-Learning, 15*(5), 369–383. https://academic-publishing.org/index.php/ejel/article/view/1845

Fermín-González, M. (2019). Research on virtual education, inclusion, and diversity: A systematic review of scientific publications (2007–2017). *International Review of Research in Open and Distance Learning, 20*(5), 146–167. https://doi.org/10.19173/irrodl.v20i5.4349

González, C. (2010). What do university teachers think eLearning is good for in their teaching? *Studies in*

Higher Education, 35(1), 61–78. https://doi.org/10.1080/03075070902874632

Graf, S., Liu, T.-C., & Kinshuk. (2010). Analysis of learners' navigational behaviour and their learning styles in an online course. *Journal of Computer Assisted Learning, 26*(2), 116–131. https://doi.org/10.1111/j.1365-2729.2009.00336.x

Gürcan, F. (2009). *Web içerik madenciliği ve konu sınıflandırılması.* Karadeniz Teknik Üniversitesi. http://acikerisim.ktu.edu.tr/jspui/handle/123456789/437

Gurcan, F. (2018). Multi-class classification of Turkish texts with machine learning algorithms. *ISMSIT 2018—2nd International Symposium on Multidisciplinary Studies and Innovative Technologies.* https://doi.org/10.1109/ISMSIT.2018.8567307

Gurcan, F. (2019). Extraction of core competencies for big data: Implications for competency-based engineering education. *International Journal of Engineering Education, 35*(4), 1110–1115. https://www.ijee.ie/contents/c350419.html

Gurcan, F., & Cagiltay, N. E. (2020). Research trends on distance learning: A text mining-based literature review from 2008 to 2018. *Interactive Learning Environments.* https://doi.org/10.1080/10494820.2020.1815795

Gurcan, F., Cagiltay, N. E., & Cagiltay, K. (2021). Mapping human–computer interaction research themes and trends from its existence to today: A topic modeling-based review of past 60 years. *International Journal of Human–Computer Interaction, 37*(3), 267–280. https://doi.org/10.1080/10447318.2020.1819668

Hung, J. L. (2012). Trends of e-learning research from 2000 to 2008: Use of text mining and bibliometrics. *British Journal of Educational Technology, 43*(1), 5–16. https://doi.org/10.1111/j.1467-8535.2010.01144.x

Kaizer, B. M., Sanches da Silva, C. E., Zerbini, T., & Paiva, A. P. (2020). E-learning training in work corporations: A review on instructional planning. *European Journal of Training and Development, 44*(6/7), 615-636. https://doi.org/10.1108/EJTD-08-2019-0149

Khanal, S. S., Prasad, P. W. C., Alsadoon, A., & Maag, A. (2020). A systematic review: Machine learning based recommendation systems for e-learning. *Education and Information Technologies, 25*, 2635–2664. https://doi.org/10.1007/s10639-019-10063-9

Kibuku, R. N., Ochieng, D. O., & Wausi, A. N. (2020). E-learning challenges faced by universities in Kenya: A literature review. *Electronic Journal of e-Learning, 18*(2), 150–161. https://doi.org/10.34190/EJEL.20.18.2.004

Klingenberg, O. G., Holkesvik, A. H., & Augestad, L. B. (2020). Digital learning in mathematics for students with severe visual impairment: A systematic review. *British Journal of Visual Impairment, 38*(1), 38–57. https://doi.org/10.1177/0264619619876975

Krull, G., & Duart, J. M. (2017). Research trends in mobile learning in higher education: A systematic review of articles (2011–2015). *International Review of Research in Open and Distance Learning, 18*(7). https://doi.org/10.19173/irrodl.v18i7.2893

Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics, 106*, 213–228. https://doi.org/10.1007/s11192-015-1765-5

Patra, S. K., Bhattacharya, P., & Verma, N. (2006). Bibliometric study of literature on bibliometrics. *DESIDOC Journal of Library & Information Technology, 26*(1). https://doi.org/10.14429/djlit.26.1.3672

Rodrigues, H., Almeida, F., Figueiredo, V., & Lopes, S. L. (2019). Tracking e-learning through published papers: A systematic review. *Computers and Education, 136*, 87–98. https://doi.org/10.1016/j.compedu.2019.03.007

Rodrigues, M. W., Isotani, S., & Zárate, L. E. (2018). Educational data mining: A review of evaluation process in the e-learning. *Telematics and Informatics, 35*(6), 1701–1717. https://doi.org/10.1016/j.tele.2018.04.015

Rowley, J., & Slack, F. (2004). Conducting a literature review. *Management Research News, 27*(6), 31–39. https://doi.org/10.1108/01409170410784185

Tibaná-Herrera, G., Fernández-Bajón, M. T., & De Moya-Anegón, F. (2018a). Categorization of e-learning as an emerging discipline in the world publication system: A bibliometric study in SCOPUS. *International Journal of Educational Technology in Higher Education*, *15*(1), 21. https://doi.org/10.1186/s41239-018-0103-4

Tibaná-Herrera, G., Fernández-Bajón, M. T., & De Moya-Anegón, F. (2018b). Output, collaboration and impact of e-learning research: Bibliometric analysis and visualizations at the country and institutional level (Scopus 2003–2016). *Profesional de La Informacion, 27*(5), 1082–1096. https://doi.org/10.3145/epi.2018.sep.12

Valverde-Berrocoso, J., del Carmen Garrido-Arroyo, M., Burgos-Videla, C., & Morales-Cevallos, M. B. (2020). Trends in educational research about e-learning: A systematic literature review (2009–2018). *Sustainability (Switzerland), 12*(12), 5153. https://doi.org/10.3390/su12125153

Yang, X. L., Lo, D., Xia, X., Wan, Z. Y., & Sun, J. L. (2016). What security questions do developers ask? A large-scale study of stack overflow posts. *Journal of Computer Science and Technology, 31*, 910–924. https://doi.org/10.1007/s11390-016-1672-0

Zitzmann, N. U., Matthisson, L., Ohla, H., & Joda, T. (2020). Digital undergraduate education in dentistry: A systematic review. *International Journal of Environmental Research and Public Health, 17*(9), 3269. https://doi.org/10.3390/ijerph17093269

Investigation of Emerging Trends in the E-Learning Field Using Latent Dirichlet Allocation
Gurcan, Ozyurt, and Cagiltay

18