USING A TEXT MINING APPROACH TO EXAMINE ONLINE LEARNING RESEARCH TRENDS OF THE PAST 20 YEARS (1997-2016)

Heather Lynn Keahey

Dissertation Prepared for the Degree

DOCTOR OF PHILOSOPHY

UNIVERSITY OF NORTH TEXAS

December 2017

APPROVED:

Lin Lin, Major Professor
Robert Wright, Committee Member
J. Michael Spector, Committee Member
Cathleen Norris, Chair of the Department of Learning Technologies
Kinshuk, Dean of the College of Information
Victor Prybutok, Dean of the Toulouse Graduate School

Keahey, Heather Lynn. *Using a Text Mining Approach to Examine Online Learning Research Trends of the Past 20 Years (1997-2016)*. Doctor of Philosophy (Learning Technologies), December 2017, 174 pp., 19 tables, 49 figures, references, 183 titles.

The purpose of this research is to identify longitudinal trends relevant to online learning research within 15 highly regarded, peer-reviewed publications in educational technology and online education. Online instruction has become a popular form of education delivery across academic institutions. A review of literature on the topic shows that missing from the corpus is a trend analysis focused in online learning research across multiple journals. Previous efforts of establishing trends in online learning are narrow in focus using only one journal or a shortened time frame. This metatrend analysis employed text mining techniques to examine twenty years (1997-2016) of published research in an effort to establish past, present and emerging trends within published literature. A general bibliometric analysis is offered highlighting prolific and yearly journal publications. Meaningful trending terms used during the twenty-year time period were identified and analyzed. A cluster analysis performed on the extracted data provides a single layer taxonomy regarding online learning research. Time trends within the clusters were identified to offer a more in-depth analysis. Trends revealed during the research indicate a changing relationship of online learning and distance education. A strong emphasis on students and learning was noted as a consistent trend throughout the literature. Emerging categories recognized include openness and mobility, game-based learning, and MOOCs. The intention of the research is to offer an overview of trends in online learning research in order to contribute to the ongoing dialogue concerning the development and delivery of online education.

ACKNOWLEDGEMENTS

Thank you to the Learning Technologies PhD program and UNT. It was a privilege to work with the faculty and be a part of the 2018 cohort. My wonderful committee was ever vigilant and attentive. To my Major Professor, Dr. Lin Lin, my mentor and role model: I feel honored to be one of your PhD students. Thank you, Dr. Spector; your wisdom and insights always add immense value to any work. Thank you, Dr. Wright, for all the thoughtful and encouraging words throughout my entire journey. I also want to thank my incredible research partner, Gwen Morel. It is my hope that we can collaborate on many projects for years to come.

Throughout this process, I have had unwavering support of many, many friends and especially my family. Thank you to my mom and dad, Dr. John and Patty Johnson, for always loving and supporting me. You truly are the greatest parents. Thank you, Mary Sue & Eldred Keahey, the greatest in-laws, for your optimism and support. To my wonderful sisters and brothers and sisters-in-law: Holly and Andy, Harmony and Ben, Hamp and Marilee, Andria and John, Theresa and Jerry, thank you for your constant support and encouragement. All of you are an inspiration to me. To my awesome nieces and nephews (27 in all), I wish for you to always dream big (A special shout out to Xan for being one of my many readers.). To my children: Lindsay, Kimberly & Matt, and Mindy, I am so proud of each of you. You are amazing, and I feel tremendously blessed to be your mom. Thank you for all your help and constant encouragement. For my husband Kelly, words cannot express my gratitude for the love and joy you have given me the past 26 years. Throughout this journey, you have been my rock and my biggest cheerleader. You are a constant reminder that:

I can do all things through Christ who strengthens me.
Philippians 4:13

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

The increasing offerings of online programs within education necessitate a continued focus on the design and delivery of effective online instruction (Trespalacious & Rand, 2015). Online learning options address several issues of stakeholders within education calling for school choice and reform, as well as teaching of 21st century skills that prepare students for the workforce (Cavanaugh, Gillan, Kromrey, Hess, & Blomeyer, 2004). Web-based education at one time could be considered, at best, an evolving technology tool and, at worst, an inferior delivery mechanism for instruction. However, over time, online learning has garnered support from mainstream educational institutions. As recent as 2013, the perception that online learning was comparable to traditional classroom instruction was decidedly mixed; though, over 75% of leaders in academia believed that online education is as good as or in some cases better than a face-to-face format (Allen & Seaman, 2013). In a broad sense, online learning can and should influence accessibility ensuring a student's right to an education (Lin, 2008). It can be said that a new norm has emerged about opening access to educational opportunities through online education (Naidu, 2014).

Statement of the Problem

Since the Internet became a universally accessible entity, educational opportunities with new pedagogical models have emerged (Harasim, 2000). Examples include the expansion of multimedia content and implementation of online discussion forums. These and other changes following the advent of the Internet have altered the landscape of the distance and open education fields. For effective research in the field of online and open education to be accepted and used, it is important for researchers to be aware of trends in the field (Wong, Zeng, & Ho, 2016). To properly identify the trends, peer-reviewed academic research documents concerning online learning should be examined. A *trend* can be described as a prevailing tendency or a general movement (trend, 2016). Educational institutions may not control the directions of trends in technology but could leverage some of the trends to improve both teaching and learning (Natividad, 2016).

Over the past two decades, an abundance of academic research has been published related to online education. The problem is that a study identifying longitudinal trends within online learning research across multiple journals does not exist. Additionally, due to the vast number of academic articles published on the subject of online learning, trying to examine the literature would be taxing in terms of time and resources. This study addresses the profusion problem using text-mining, an automated process for analyzing text-based digital data. Text mining can be used to help overcome the hurdles associated with examining large amounts of textual-based documents (Delen & Crossland, 2008). Articles from selected journals within the fields of educational technology and online education will serve as the corpus of literature for the study.

Purpose of the Study

Delivery of online education has changed considerably during the past 20 years. Examining the trends within the scholarly research could provide valuable information to scholars in the area of online learning. By analyzing the content of articles published in highly regarded educational technology and online or distance education journals, trends within the research can be identified. This information can give guidance to researchers and practitioners by providing an overview on past and current status of research in online learning.

Publication in peer-reviewed academic journals has remained the standard way of disseminating scientific research. With the progression of Web-based technology, delivery of information and communication among researchers and practitioners is becoming easier and faster. This can be seen with the increasing presence of e-journals such as *Online Learning* and the *European Journal of Open, Distance, and E-Learning.* However, if publication procedures bypass quality control mechanisms, they risk degrading the standards that dictate acceptable scientific research (Bontis & Serenko, 2009). For this reason, each journal chosen for inclusion in this study has been evaluated for its appropriateness and contribution to educational technology and online learning.

The purpose of this study is to provide a metatrend analysis by examining trends within online learning literature over a 20-year time period (1997-2016). Fifteen highly regarded journals in the fields of educational technology and online/distance education have been identified for use in this study. This research is an extension of a prior effort investigating the trends within educational technology research literature (Natividad, 2016). The current study further examines one of the noted trends in that research, online learning. The body of online learning literature in this study exceeds 4,000 articles. An automated process of analyzing the articles through the use of text-mining techniques can provide an objective analysis of the vast amount of research literature. This study examines the longitudinal trends within the literature and offers a categorical taxonomy based on thematic clusters found within the text.

Research Questions

The research questions for this study are:

1. What are the bibliometrics of online learning articles among the 15 journals during the past 20 years (1997-2016)?

2. What are the past, present, and emerging trends of terms regarding online learning within the selected literature during the past 20 years (1997-2016)?
3. What are the thematic clusters of the articles regarding online learning in the 15 journals for the past 20 years (1997-2016)?
4. What are the time trends of the recognized thematic clusters found in the selected articles during the past 20 years (1997-2016)?

The four research questions intentionally build upon one another with the outcome of each query laying a basis for the subsequent research questions. The first question concerning the bibliometrics of the articles provides a foundation for the entire study by identifying general and time-sensitive publication information regarding online learning research. The second question about the trending terms produces data regarding prominent terminology used within the research. The aggregation of the terms serves as input data to identify clusters within the published research corpus. The last research question addresses issues involving the historical timeline of the identified clustered categories.

Rationale

The impetus for the proposed research stems from prior published work. In 2016, Natividad (2016), as part of her dissertation, published results from her study that examined research articles published in ten highly regarded educational technology journals. The title of her work is: *An Analysis of Educational Technology Publications: Who, What and Where in the Last 20 Years.* Natividad's (2016) research focused on concepts in educational technology, both growing and diminishing for the years 1995-2014. The results informed researchers and practitioners as to trends in the field of educational technology. An examination of the literature with a focus towards online learning could potentially yield even more useful data considering the growth and changes in this area of education delivery. The following sections discuss the selection of the original 10 journals and the additional journals, which focus on concepts related to online learning.

Selection of the 10 Journals

Natividad (2016) chose 10 journals that were identified as top journals in the field of educational technology. These journals are included in the current research. Dr. J. Michael Spector in collaboration with journal editors associated with the National Technology Leadership Coalition (NTLC) identified the key journals to be used in research examining trends within educational technology (Natividad, 2016). Since Natividad's (2016) study looked at published literature for 20 years, the journals boasted a publication history spanning two decades in order to have meaningful contributions to the research effort . During this 2-year process, five distinct criteria were used in evaluative discussions about each potential journal (Natividad, 2016):

- Impact factor: A 5-year impact factor was considered an important consideration but was not used to exclude a publication that had significant influence in the field.
- Scope: The scope of a journal should not be too narrowly focused on one particular dimension of educational technology or too broad as to encompass other aspects of education.
- Focus: The journal should emphasize research findings rather than product reports or anecdotal discussions.
- Readership: The readership of the journal should represent the field of educational technology on a global scale rather than locally or regionally.
- Authorship: Authorship should be open worldwide to researchers and not exclusively those affiliated with a specific group.

Many of the journals were excluded as they failed to meet two or more of the conditions. However, the group of editors unanimously agreed that *Educational Technology*, while not peer-reviewed or indexed, should be included in the selected group. The decision was based on the fact that the publication prints high-quality research articles that are widely read. In fact, this magazine is more widely read than the other included publications (Natividad, 2016).

The 10 educational technology journals that were used in Natividad's (2016) research and also used in this study are:

- *British Journal of Educational Technology* (*BJET*)
- *Computers and Education*
- *Educational Technology: The Magazine for Managers of Change in Education*
- *Educational Technology Research and Development (ETR&D)*
- *Instructional Science*
- *Journal of Educational Computing Research*
- *Journal of Educational Technology & Society*
- *Journal of the Learning Sciences*
- *Journal of Research on Technology in Education*
- *TechTrends*

One publication, *Educational Technology: The Magazine for Managers of Change in Education*, has ceased publication. Considered a pioneer periodical in the field of educational technology, the magazine commenced publishing in the 1960s (Educational Technology Magazine, n d.). The last year of publication was 2016. Another respected journal, the *Journal of Educational Technology & Society*, will cease publishing soon as the editors have stopped accepting new submissions (Kinshuk, Chen, & Sampson, 2016). The ending of these long-standing contributors to the field of educational technology adds value considering the timing of this study.

Selection of the Five Additional Journals

As this study's focus was specific to online learning, additional journals specializing in the field were selected to be included in the data set. By adding additional research from online education-focused journals, a more complete picture of research trends on the topic can be developed. A group of three experts in the field of Learning Technologies along with a PhD candidate from the University of North Texas's Department of Learning Technologies, comprised a list of 17 widely read journals dedicated to distance and/or online learning:

- *American Journal of Distance Education*
- *Distance Education*
- *E-Learning and Digital Media*
- *E-Learning and Education*
- *Electronic Journal of e-Learning*
- *European Journal of Open, Distance, and E-Learning.*
- *International Journal of Distance Education Technologies*
- *International Journal of E-Learning*
- *International Journal of E-Learning and Distance Education*
- *International Journal of Instructional Technology and Distance Learning*
- *International Journal of Online Pedagogy and Course Design*
- *International Review of Research in Open and Distance Learning*
- *Journal of Interactive Online Learning*
- *Online Journal of Distance Learning Administration*
- *Online Learning Journal (Formerly Journal of Asynchronous Learning)*
- *Open Learning: The Journal of Open, Distance, and e-Learning*
- *Quarterly Review of Distance Education*

To maintain consistency within the corpus of selected literature, the journals were vetted on the basis of publication history and impact factor. For an additional journal to be used in the study, it must have maintained a 20-year publication history and have an impact factor comparable with the previous 10 selected journals. Five of the 17 journals met these criteria and were selected for this study. These journals are:

- *American Journal of Distance Education*
- *Distance Education*
- *European Journal of Open, Distance, and E-Learning.*
- *Online Learning Journal (Formerly Journal of Asynchronous Learning)*
- *Open Learning: The Journal of Open, Distance, and e-Learning*

The remaining 12 journals have a publication history less than 20 years, with the exception of one journal. *The International Journal of E-Learning and Distance Education* has a publication length greater than 20 years; however, a reliable impact factor calculation measuring article citation was not located to support inclusion in this study. Details concerning the publications history of these 17 journals can be found in Appendix B.

Synopsis of the 15 Journals

Journal covers for the selected periodicals are shown in Figure 1. The first 10 journals are from Natividad's (2016) original research. The last five journals, which are shown with a different border, were added to provide perspective from journals dedicated to online and distance education. These 15 journals contain articles providing the data used in this study. Each journal is highly regarded and maintains a positive influence in the educational technology arena.

Figure 1. The cover pages of the 15 journals included in this study, which includes 10 educational technology and five distance education publications.

One method for measuring impact of an academic journal is through an h-index calculation. The h-index measures the scientific productivity and influence in the academic community based on the number of citations from a journal over a period of time (National Institutes of Health, 2017). The measurement accounts for the number of papers published in a journal along with the number of citations. A complete list of journals for this study along with an h-index, according to the SCImago Journal Rank (SCImago, 2007) is given in Table 1. For two of the additional journals, an index in this ranking could not be located. An h-index calculation found in the Google Scholar Metric page (2017) is shown for these journals.

Table 1

List of Included Journals H-Index Measure from SCImago Journal Rank

Journal Title	H-Index
British Journal of Educational Technology	63
Computers & Education	109
Educational Technology Magazine	n/a
Educational Technology Research and Development (ETR&D)	63
Instructional Science	51
Journal of Educational Computing Research	42
Journal of Educational Technology & Society	55
Journal of the Learning Sciences	70
Journal of Research on Technology in Education	6
TechTrends	22
The American Journal of Distance Education	8
Distance Education	34
European Journal of Open, Distance, and E-Learning	13*
OLC Online Learning Journal	29
Open Learning: The Journal of Open, Distance, and e-Learning	13*

Note. *Denotes h-index retrieved from Google Scholar.

Each journal in this study brings a unique contribution to the educational technology and online learning fields. Publication figures for each journal is provided in Table 2. Interesting to note is that the original 10 journals typically publish more issues per year than the additional five online/distance learning journals. Also, not surprisingly, is that these 10 journals generally contain more published articles over the twenty-year period.

Table 2

Journal Publication Information

Journal Name	Issues per Year*	Total Issues (1997-2016)	Total Articles (1997-2016)	Average Articles per issue
British Journal of Educational Technology	6	109	1472	13.50
Computers and Education	12	172	2666	15.50
Educational Technology Magazine	6	120	986	8.22
Education Technology Research and Development	6	102	725	7.11
Instructional Science	6	119	570	4.79
Journal of Educational Computing Research	8	95	865	9.11
Journal of Educational Technology & Society	4	73	1467	20.10
Journal of the Learning Sciences	4	80	272	3.40
Journal of Research on Technology in Education	4	78	456	5.85
Tech Trends	6	120	921	7.68
American Journal of Distance Education	4	75	298	3.97
Distance Education	3	52	406	7.81
European Journal of Open, Distance, and E-Learning	2	34	338	9.94
OLC Online Learning Journal	4	67	532	7.94
Open Learning: The Journal of Open, Distance, and e-Learning	3	60	382	6.37
Totals		1,356	12,356	

* Issue numbers based on 2016 data

The original 10 journals are recognized for their impact and contributions in educational technology. Many of these journals maintain a substantial focus on research in educational technology: (a) *British Journal of Educational Technology*, (b) *Computers and Education*, (c) *Education Technology Research and Development*, (d) *Journal of Educational Computing Research,* and (e) *Journal of Research on Technology in Education*. A couple of the journals, *Instructional Science* and *Journal of the Learning Sciences*, have a broader research emphasis in education. Two other publications support practitioners and those who develop and implement systems in the field of educational technology: *Journal of Educational Technology & Society* and *TechTrends*. Within the group of 10 is a magazine, *Educational Technology,* though not indexed as an academic journal, has had considerable influence in the field. These 10 journals as part of their subject matter, contain articles related to online learning.

Five additional journals, while focused in online/distance education, each offer a unique contribution. The *American Journal of Distance Education* is described as a journal of record in distance education for the Americas (AJDE, 2017). Another journal with a continental concentration is the *European Journal of Open, Distance, and E-Learning* which, as it name indicates, houses material of European interests relating the three educational issues of openness, distance and e-learning (European Distance and E-Learning Network, n.d.). Another journal boasting the same focus trilogy and with a global interest is *Open Learning* (Open University, 2017). Based in Australia, *Distance Education* covers topics related to open, distance and flexible learning (Open and Distance Learning Association of Australia, Inc., 2017). Sponsored by the Online Learning Consortium, *Online Learning,* is dedicated to publishing articles related to online learning (Online Learning Consortium, 2017). Though some of the journals are associated with a geographical region, all contain articles from authors worldwide. The readership and applications within these journals maintain a global reach.

Research Methods

All articles contained in theses 15 journals from 1997-2016 were considered for use in this study. Articles selected from these journals contain a focus in online learning. The titles and abstracts of all articles in the 15 journals were searched for text strings of terms related to online learning. If an abstract or title contained one of the stringed terms, the article was flagged. Each selected article was reviewed to determine the appropriateness for the study. From the articles designated for use in the online learning research effort, a bibliometric analysis was performed. Bibliometric measures allow researchers to generate quantitative information from large amounts of historical text-based documents; however, a criticism of the process is that it focuses on numbers and not actual content of the documents represented (Hung, 2012). The bibliometric analysis for this study is based on publication years and source journals in an effort to provide a foundation for establishing the trends.

Text mining was used to glean meaningful information regarding the content of the selected articles. Automated text mining using Rapid Miner (2017) was employed to extract terms from the abstracts. The list of terms was assembled with inconsequential terms, symbols, and spaces removed from the data set. Once the process was completed, each unique term was totaled. To provide a snapshot of trending terms throughout the 20 years, five separate summation calculations occurred. Four of the summations are for 5-year distinct time periods: 1997-2001, 2002-2006, 2007-2011, and 2012-2016. Additionally, a tally was performed on the entire text data for the 20-year span (1997-2016). The initial output of the largest term extraction is used for the next process, a cluster analysis.

Once the terms were extracted and filtered, a cluster analysis was performed. A k-means cluster analysis was used to separate the articles into thematic clusters. This type of analysis is considered a form of Latent Semantic Analysis (LSA) (Evangelopoulos, Zhang, & Prybutok, 2012). In a k-means clustering algorithm, the weighted text data is separated into categories or clusters. The *k* value, which determines the number of clusters, was inferred from a data plot of group of sum squares against the number of clusters. Once the clusters were formed, they were labeled according to the highest weighted terms and cluster contents. Data regarding publication years within the clusters was evaluated to provide categorical time trend information.

Operational Definitions

Bibliometrics

Bibliometric analysis provides a quantitative measure of scientific literature by certain indicators (Thelwall, 2008). Keshava, Gireesh, and Gowda (2008) describe it as a summation of publication information with quantitative statistics regarding growth of papers by year, ranking of prolific information, and other descriptors. This study uses a basic form of bibliometrics providing yearly publication statistics and basic prolific journal measures.

Centroid

A centroid is defined as a center of mass (centroid, 2017). When finding the center of a group of text, there is no tangible mass. Within data, it is considered a point (imaginary or real) at the center of a calculated cluster (Söder, 2008). In this study, a k-means clustering algorithm is used to identify clusters from textual data. The centroid, or mathematical mean, is the point in the center around which each cluster is formed.

Cluster Analysis

A cluster analysis divides a collection of text-based documents into unique groups or clusters based on themes found within each document (Chakraborty, Pagolu, & Garla, 2014). It has widespread use in studies involving text mining (Delen & Crossland, 2008). A k-means cluster analysis, which produces a single level of groupings, was chosen for categorizing the data in this study. One of the simplest clustering algorithms is the k-means analysis, which has long history of use in various scientific fields (Jain, 2010). Though other clustering processes exists, the k-means program proved to be suitable and accessible. Partitional clustering algorithms separate the given *n* data sets into *k* partitions, where each represents a subset or a cluster (Prabha, Duraiswamy, & Sharmila, 2016). The goal is to define *k* centroids, or centers, one for each respective cluster ("Clustering: An introduction", n. d.) This research employs a k-means algorithm and the term cluster analysis will refer to this specific process unless otherwise noted.

Distance Education

In a publication sponsored by the Association for Educational Communications & Technology (AECT), distance education is defined as "institution-based, formal education where the learning group is separated, and where interactive telecommunications systems are used to connect learners, resources, and instructors." (Schlosser & Simonson, 2003, para. 1). The delivery of educational content does not depend on the geographical location of the student or instructor. By acknowledging physical separation of teacher and pupils, distance education does not maintain a fixed place or fixed time for training or education (Keegan, 1995).

E-Learning

The term e-learning can be traced to the mid-1990s as development and interest in the World Wide Web grew (D. R. Garrison, 2011). The term has a close association with online learning. E-learning has been defined as the delivery of instructional content to individuals using digital network technology (Welsch, Wanberg, Brown, & Simmering, 2003). However, e-learning can represent other forms of electronic instruction that do not require Internet technology to operate (J. L. Moore, Dickson-Deane, & Galyen, 2011). This research document follows most closely with the definition given by Clark and Mayer (2016). They define e-learning as:

> Instruction delivered on a digital device that is intended to support learning. In e-learning the delivery hardware can range from desktop or laptop computers to tablets or smart phones, but the instructional goal is to support individual learning or organizational performance goals. (p. 7)

This definition of e-learning does not strictly refer to Internet hardware technologies as the devices listed can run on a network or as a stand-alone device without any network ties.

Latent Semantic Analysis

The fundamental idea behind Latent Semantic Analysis (LSA) can be described as:

> The meaning of each passage of text (*a document*) is related to patterns of presence or absence of individual words, whereas a collection of documents (*a corpus*) is modeled as a system of simultaneous equations that can determine the similarity of meaning or words and documents to each other (Evangelopoulos et al., 2012, p. 71).

This study considers cluster analysis performed with data retrieved from an unstructured text-based document as a viable practice of LSA. This is in alignment with other researchers within the text-mining community (Antai, Fox, & Kruschwitz, 2011; Evangelopoulos et al., 2012;

Kireyev, 2008). The current study follows the stated definition and uses LSA as an umbrella under which a cluster analysis of text-based data suitably fits.

Meta–Analysis, Trend Analysis, and Metatrend Analysis

A *meta-analysis* combines data from multiple studies (Biostat, 2017) and is a recognized practice for synthesizing research outcomes to provide a basis for understanding a phenomenon and a "parsing of influences on the phenomenon" (Cavanaugh et al, 2004, p. 8). The aim of a meta-analysis is generally to discover agreement or disagreement within the results of similar studies in order to scale outcomes to a broader population than typically exists in one focused study. A meta-analysis research approach is widely used in a variety of fields including, but not limited to, education, medicine, criminal justice, political science, and sociology (Biostat, 2017)

A *trend analysis* collects information in order to recognize a pattern or relationship between associated factors or variables (trend analysis, 2017). The information collected can be in a variety of formats. One common way is to collect survey information as Immerwahr (2004) did to produce a trend analysis on public attitudes towards higher education. The data for a trend analysis can, but many times does not, originate from academic research.

For purposes of the current research, a *metatrend analysis* is a process in which the aim is to identify trends within peer-reviewed academic literature. Metatrend analyses have been published previously within educational technology. One specific example is Hung and Zhang's (2012) examination of longitudinal trends in their article titled, *Examining mobile learning trends 2003-2008: A categorical meta-trend analysis using text mining techniques*. However, an earlier metatrend analyses is found in the research journal, *Pain*, where categorical trends from 32 years of published research in the journal were identified and analyzed (Mogil, Simmonds, &

Simmonds, 2009). Other metatrend analyses have since been published within several different research fields including medicine (Coronado, Riddle, Wurtzel, & George, 2011), educational technology (Hung, 2012), music and cognition research, (Tirovolas & Levitin, 2011), and non-profit organizations (Goddard & Annaraj, 2017).

MOOCs

Massive open online courses (MOOCs) are, as the name implies, large enrollment classes delivered using Internet technology. The courses tend to involve hundreds or thousands of students making the scale rather relative (Siemens, 2013). MOOCs are usually open source online courses that integrate "the connectivity of social networking, the facilitation of an acknowledged expert in a field of study, and a collection of freely accessible online resources" (McAuley, Stewart, Siemens, & Cormier, 2010, p. 4). Like typical academic courses, MOOCs have a designated set start and end time (Siemens, 2013).

N-Gram

An n-gram is defined as a contiguous sequence of a number of items (n) from a certain string of text or speech (n-gram, n. d.). The items can be characters, words, or other specified elements. Used extensively in text mining, they are a set of words occurring consecutively within a given piece of text (Ganesan, 2014). For this research, the items in the n-grams are words.

OER

The term *OER* is used in educational research as an acronym for open educational resources. This term was defined by the United Nations Educational Scientific and Cultural Organization (UNESCO) in 2002 and was further defined in 2004 to include learning resources, teacher support resources, and quality assurance resources (Johnstone, 2005). Open education is based on the sharing of open educational resources across demographics which remove economic barriers (de Langen & Bitter-Rijkema, 2012). The definition adopted for this work is from Geith and Vignare (2008), who state that OER "includes free (no charge) and open (for modification) resources such as digital content, open source software, and intellectual property licenses" (p. 106).

Online Education

Education can be defined as a "process designed to help others learn" (Rogers, 2003, p. 5). The term online education implies that delivery of instruction to the student is through a connection to a computer system or device (Volery & Lord, 2000). For purposes of this research online education signals the process by which instructional content is transported to the learner.

Online Learning

By its nature, online learning is difficult to define (J. L. Moore et al., 2011). J. F. Watson and Kalmon (2005) describe *online learning* as "education in which instruction and content are delivered primarily via the Internet" (p. 127). Another definition offered in the literature is that "online learning is a form of distance education where technology mediates the learning process, teaching is delivered completely using the Internet, and students and instructors are not required to be available at the same time and place" (Joksimović, et al., 2015, p. 100). Other authors recognize online learning as a more recent version of distance learning (Benson, 2002; Carliner, 2004; Conrad, 2002). The term is commonly used interchangeably with cyber learning, virtual learning, and e-learning (International Association for K-12 Online Learning (iNACOL), 2011). Debate exists for the proper use of the terms within the learning technology community. Cyber-learning, though becoming arguably outdated, can be used interchangeable with online learning (International Association for K-12 Online Learning (iNACOL), 2011). Online learning can be considered as virtual learning. However, the use of virtual learning is in many cases associated with a more specific connotation, such as virtual learning environments (J. L. Moore et al., 2011; Paulsen, 2002) or virtual learning communities (Matzat, 2013). As mentioned previously, e-learning can represent other forms of electronic instructional delivery. For purposes of this study, the definition of online learning acknowledges the connection with distance education and also indicates the online delivery of educational content.

Text Mining

Text mining is the process of extracting meaningful patterns from unstructured text files and is considered an extension of data mining (Feldman & Dagan, 1995). The process depends on computational analysis of the textual data and uses domain experts for evaluation and interpretation (Hung, 2012). Applications for text mining include document classification and clustering, information extraction, natural language processing, concept extraction, Web mining, and information retrieval (Miner, Delen, Fast, Hill, & Nisbet, 2012). In this study, text mining refers to the extraction of meaningful terms and the classification of text-based documents by an automated process using computer software.

Limitations

The author acknowledges and recognizes that limitations are present within this body of research. Placing a limit on the number of journals accepted in this study, while necessary for the integrity and practicality of the study, most likely will exclude publications that could potentially make a valuable contribution to this work. The criteria for selection of journals in this study ensured the inclusion of historically, highly regarded publications in the fields of educational technology and online education. Journals that maintained a limited publication history were excluded due to the longitudinal nature of this study. Many journals that focus on online education do not have a 20-year publication history. Journals without a 20-year publication history and other sources such as conference proceedings or reputable Websites could provide a depth of research or a valid perspective that is not represented in the results of this study.

Another limitation noted is with the k-means clustering process being used to classify the articles. This method is a popular and widely used process in text mining. However, other organizational algorithms exist, such as hierarchical clustering and factor analysis, which could provide a different dimension in the organizational structure. It should also be noted that there are other more specialized forms of k-means clustering that were not used in this study.

One further limitation is that there is a qualitative aspect to this research, which by its nature brings a bias into the process (Roulston & Shelton, 2015). The interpretation and labeling of the clusters was done by experts in the field of learning technologies but is subject to any bias these experts bring to the process. Effort was made to provide a fair evaluation and appropriate labels for the research categories.

While limitations within this study are recognized, it does not prevent the offering of meaningful longitudinal trend analysis of published online learning research. Where possible, objective standards were used in the selection process of the articles and journals, the text mining of the article abstracts, the labeling of the clusters, and the interpretation of the results. This research is intended to provide a preliminary overview of the trends found in online learning research for the past 20 years.

CHAPTER 2

REVIEW OF LITERATURE

The purpose of this literature review is to examine published research centered in online learning, seeking to give a broad overview of findings. The findings from the review provide background information in addressing the research questions. Four research questions give an overarching guidance towards this study. These questions concern the (a) bibliometrics, (b) term trends, (c) categories, and (d) categorical time trends of research about online learning.

As online learning enrollment continues to grow (Straumsheim, 2017) up-to-date research on the topic becomes imperative especially when considering the continual advances in technology. Online learning research efforts often include the broader subject of distance education, since the origins of online learning can be traced back to this concept (Taylor, 2001). However, online learning has become more commonplace and its implementation goes beyond its role in distance education. Current research should reflect this broader application. Two popular topics among meta-analysis research focus on discussions of the effectiveness of distance or online education and evidences of trends within the subject. Presented in this chapter are three sections. First, a history of online learning and its relationship with distance education is discussed. The subsequent section contains a review of meta-analysis literature focused on online or distance education topics. The third section references automated research techniques (text mining) used to conduct analyses relevant to the topic of online learning.

Online Learning and Distance Education

Within the body of literature, online education has been historically closely linked with distance education. Online learning intersects with the larger and more established category of distance education, which itself includes correspondence courses, educational television, and videoconferencing (Means, Toyama, Murphy, & Baki, 2013). Early attempts to offer online education tended to mimic the practices found in existing distance education (Joksimović, et al., 2015). A historical, temporal classification of distance education delivery was given by Taylor (2001), offering five distinct generations. The first four are: (a) correspondence, (b) multimedia, (c) tele-learning, and (d) flexible learning. The most recent fifth generation includes online interactive multimedia, Internet-based access to resources, computer-mediated communication, and “campus portal access to institutional processes and resources” (Taylor, 2001, p. 3).

Online and distance education maintain similar characteristics, including the emphasis that education be delivered at any time, any place, with the assumption that students and the instructor are physically separated (M. G. Moore, 1993). Anderson (2009) contends that the most compelling feature of the online medium is “the capacity of shifting the time and place of the educational interaction” (p. 344). However, online education has progressed differently from previous models of distance education. Historically, distance education was based on “the ideal of autonomy and the industrial production of prepackaged study materials” (D. R. Garrison, 2011, p. 2). Online education has evolved due to the increased capacity and speed of technology and pedagogical developments and now includes more collaborative tools and approaches to learning in Web-based courses (Joksimović, et al., 2015).

The connection between distance and online education can be characterized by exploring the terminology trends within published research. The past, trending vocabulary relationships associated with distance education and online learning are shown in Figure 2. The data included in the graph was retrieved from Scopus, the largest database containing abstracts and citations of peer-reviewed literature indexing scientific journals, books and conference proceedings

(Elsevier, 2016). The articles used for the aggregation are from a distribution of studies regarding online learning indexed in the database (Joksimović, et al., 2015). According to the graphic, the number of times the phrase *online learning* was used in publications surpassed *distance education* in 2005. While the term e-learning dominated terms used post-2005, online learning generally remains the second most prominent keyword in the literature to date.

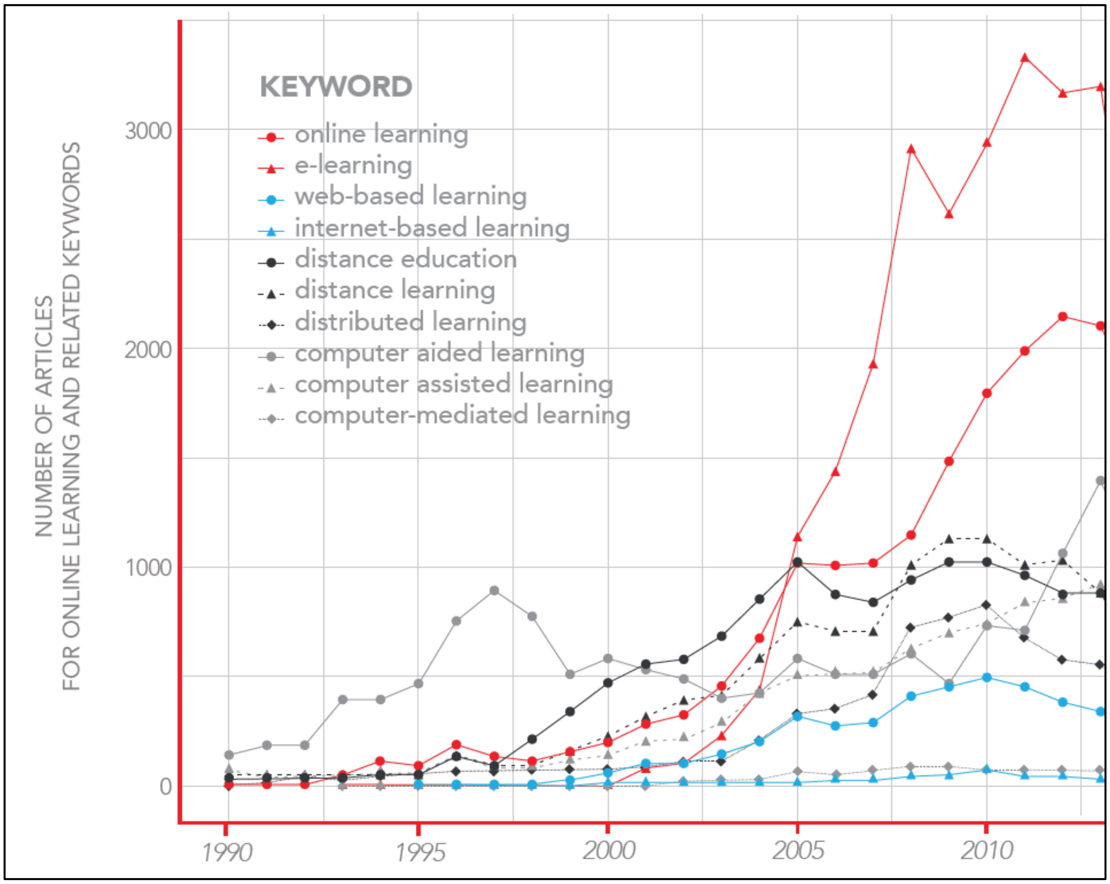

Figure 2. Comparison of terms related to online learning in research articles. Adapted from Joksimović, et al. (2015, p. 100).

Corresponding with this tendency, use of the term *online learning* has trended heavily upwards in elementary, secondary, and post-secondary education research and practice, being used as a way to offer courses at a distance (Means et al, 2013). During the 2002-2003 school year, 36% of the K-12 school districts in the United States offered online courses (Archambault

& Crippen, 2009). Ten years later, 83% of high schools reported offering online courses for their students (Bolkan, 2014). In the fall term of 2006, nearly 3.5 million students in the United States enrolled in at least one post-secondary online course (Allen & Seaman, 2007). Eight years later, in the fall of 2014, that number had grown to 5.8 million students, a 66% increase (Freidman, 2016).

While an increase in online offerings is evident, there are studies that address its impact on learning. Prior studies indicate that, when compared to traditional classroom instruction, distance education can have a similar impact on academic outcomes (Cavanaugh et al., 2004; Simonson, Schlosser, & Orellana, 2011). Although strengths can be ascertained in varied content delivery forms—including distance education, conventional classroom, online learning, and blended learning—the worth of each should be measured in its own unique context (Cavanaugh et al., 2004). As early as 2004, it was recognized that virtual schools, using rapidly advancing, Web-based technology, could provide the online student with an education as good as one received in a traditional classroom environment (Cavanaugh et al., 2004). Simonson et al., (2011) stated that "*it is not different education, it is distance education*; what is known about effectiveness in education is more often also applicable to distance education" (p. 124). The general consensus found in the reviewed distance education articles is that, when compared with a traditional classroom setting, distance education showed no significant differences in effectiveness.

Review of Meta-Analysis Literature

A meta-analysis examines published primary research and provides a second level of enquiry. By examining several secondary topically related studies, a broad overview of online learning research can be gleaned. This section presents such a tertiary, or third level, study of online learning research. While reviewing meta-analyses of online learning literature, two prominent categories were revealed and are presented in this section. The first is a comparative analysis that generally compares the effectiveness of online instructions with another mode of delivery, often, a traditional classroom setting. The second type of meta-analysis recognized in this review was a content analysis that looks for trends and categories among the articles in the larger body of literature.

Comparative Studies

Effectiveness of online learning. Debuting in the 1980s (Hickey, 2014), online learning has a relatively brief past from which to draw established research histories. Much of the reviewed literature consisted of comparative studies that examined student learning outcomes in an online setting that were compared with traditional classroom settings. Collecting these results through a meta-analysis can give a broader more consistent message regarding these outcomes. A meta-analysis published in 2004 examined students in online K-12 distance education courses (Cavanaugh et al., 2004). This quantitative synthesis effort was comprised of 14 studies and compared the effects of Web-delivered education with classroom-based settings (Cavanaugh et al., 2004). The study found online education could produce similar learning outcomes as education in a traditional classroom environment (Cavanaugh et al., 2004).

A meta-analysis of distance education research conducted in the early years of online learning included 232 studies published during an 18-year period between 1985 and 2002 (Bernard, et al., 2004). Articles reviewed by the authors included studies that spanned educational platforms, including elementary, secondary, and post-secondary students (Bernard, et al., 2004). The results suggested that some distance education implementations outperformed comparable classroom settings; some did not (Bernard, et al., 2004). Interestingly, the mean achievement effect size for synchronous distance education platforms favored a classroom setting, while asynchronous applications favored distance education (Bernard, et al., 2004).

A later meta-analysis by Bernard and his associates focused on interaction treatments of distance-delivered courses (Bernard, et al., 2009). This comprehensive work examined 74 articles to answer questions regarding student achievement when factoring in instructional conditions that influence student interaction. Three forms of interaction were evaluated: student-teacher (ST), student-student (SS), and student-content (SC). A major finding was that designing course interaction treatments, regardless of the type of interaction, has a positive effect on student learning (Bernard, et al., 2009). With wide variability in the effect sizes, distance education was shown to have better outcomes in some circumstances; in other circumstances, distance education can have worse results than classroom instruction (Bernard, et al., 2009).

A meta-analysis published the same year from the U.S. Department of Education (DOE) examined online learning and reported that, on average, learners in online environments exhibited higher academic achievement outcomes than those in a physical classroom setting (Means, Toyama, Murphy, Bakia, & Jones, 2009). However, despite the support for online instruction, the literature included in the analysis did not back the idea that online education is a superior form of instruction. Means et al. (2009) concluded from their examination that the variations in the manner of implementation of online education do not significantly impact student learning outcomes. The report spanned 13 years (1996-2008), concentrating on K-12 settings. The authors noted that online learning allowed for an increase in learning time, when compared to classroom settings (Means et al., 2009).

A subsequent meta-analysis by Means et al. (2013) shaped the statistical synthesis of online research by comparing and contrasting learning outcomes of online and traditional classroom instruction. Overall, the analysis showed that, on average, students in an online course performed modestly better than those in a face-to-face setting (Means et al., 2013). A more significant advantage was noted in blended learning, but was not observed in studies contrasting purely online environments with face-to-face settings (Means et al., 2013). It is noteworthy that the educational delivery method that produced the greatest advantage had both an online and a face-to-face component. This seems to indicate that elements exist in both traditional classrooms and online instructional environments that are advantageous to the learner.

Instructional design implications. With much research centered on the effectiveness of online instruction, it follows that these findings lead to instructional design recommendations and implementations. It has been noted that teaching with technology is much more difficult than teaching in a traditional classroom with newer instructional technologies and approaches creating a challenge for teachers (Spector, 2007). An online environment offers different tools for instructors and students such as discussion forums. In keeping with the concept, one analysis of 32 second-order research articles sought to address teaching and learning in online educational settings (Joksimović, et al., 2015).

> Our findings further indicate that contemporary research into online learning almost univocally agrees that structured online discussions with clear guidelines and expectations, well-designed courses with interactive content and flexible deadlines, and continuous instructor involvement that includes the provision of individualized, timely and formative feedback are the more promising approaches to fostering learning in online environments. (Joksimović, et al., 2015, p. 95)

While the reviewed meta-analyses give a broad picture of research related to the discipline of distance or online education, some published meta-efforts have focused on narrower aspects of the field. C.-W. Tsai, Shen, and Fan (2013) studied empirical research articles tied to online, self-regulated learning (SRL) strategies in an online setting. The authors proposed that online students must use SRL strategies and that courses should be developed accordingly (C.-W. Tsai et al., 2013). This gives support to the idea that instructional design in an online setting should purposefully incorporate both interactive and self-regulated elements.

Research synthesis tends to present that overall student online learning experiences are as effective as, or better than, face-to-face instruction. However, some researchers noted flaws in the meta-analytic research methods and results. In response to the meta-analysis commissioned by the US Department of Education (Means et al., 2009), Jaggars and Bailey (2010) stated that:

> Therefore, while advocates argue that online learning is a promising means to increase access to college and to improve student progression through higher education programs, the Department of Education report does not present evidence that fully online delivery produces superior learning outcomes for typical college courses, particularly among low-income and academically unprepared students. There is evidence that suggests that without additional supports, online learning may even undercut progression among low-income and academically underprepared students. (p. 1)

The research findings to this point generally support the suggestion that distance or online education, when properly planned and implemented with the suitable technology or pedagogy, can be at least as effective as traditional classroom instruction (Kovanović et al., 2015). Although, it must be acknowledged that, while the general averages in the reviewed literature support this assertion, meta-analyses studied for this review stated that there was wide variability in the effect sized reported among studies (Bernard, et al., 2004; Bernard, et al., 2009). One possible reason is that online learning is a relatively recent entry into the education arena. As technology advances, Web-based approaches to education are becoming commonplace at all educational levels.

Content Analysis

When contrasted with comparable study methods, content analysis is an avenue for systematic text review and examination. This procedure involves analyzing text-based documents and can include both qualitative and quantitative methods (Elo & Kyngäs, 2008). A major benefit of content analysis stems from the systematic replicable technique of multiple sources of text into content categories (Stemler, 2001). Presented in this section are discussions of both manual and automated examination techniques for content analysis.

Manual analysis. Coomey and Stephenson (2001) conducted an early content analysis of online learning research articles, presenting snapshots of trends found between 1998 and 2000. Their study had three objectives regarding distance education. The first objective was to determine proper categories within distance education research, while the other two aims were to identify the "most important" and the "most ignored" research domains in distance education (Coomey & Stephenson, 2001). The authors indicated that areas of (a) technology, (b) increased interactivity, and (c) instructional design issues dominated the research discussions to that point (Coomey & Stephenson, 2001). These three aspects confirm a shift in distance education research toward online instruction and away from distance learning topics. The authors also advised instructional designers to give attention to the issues of learner control and support and to provide opportunities for direct learner involvement (Coomey & Stephenson, 2001).

A manual review of distance education research for the years 2000-2008 was conducted in an effort to classify research areas (Zawacki-Richter, Bäcker, & Vogt, 2009). The study examined 695 articles from five different journals. The authors concluded that distance education research is dominated by the topics of (a) computer mediated communication, (b) instructional design, (c) individual learning processes, and (d) educational technology (Zawacki-

Richter, et al., 2009). This particual study noted examples of neglected areas of distance education research as (a) innovation and change management and (b) intercultural aspects (Zawacki-Richter, et al., 2009).

A more recent manual content analysis examined trends in research on distance education that took place between 2009 and 2013 (Bozkurt, et al., 2015). Seven different journals were used as article sources and through a manual content analysis of 861 research articles, the authors identified trending themes (Bozkurt, et al., 2015). Among the research areas identified were (a) interaction and communication in learning communities, (b) learner characteristics, (c) instructional design, (d) and educational technology (Bozkurt, et al., 2015). Their research also revealed the word *learning* as a top keyword within the literature. Mobile learning and OERs were recognized as new topics (Bozkurt, et al., 2015). The authors also contended that distance education research responds quickly to emerging research trends, while still giving attention to more established topics, such as student collaboration and teacher training.

A different manual analysis involved a systematic review of synchronous online learning literature (Martin, Ahlgrim-Delzell, & Budhrani, 2017). The review examined 157 articles from a variety of journals for the years 1995-2014. Several observations were noted including those related to instructional setting and content areas. The researchers found the most common variable studied in the research was perception or attitude, followed by interaction (Martin et al. 2017). This observation parallels the secondary study of C.-W. Tsai et al. (2013) on self-regulated learning in an online environment. The top content areas in Martin et al's research were English/foreign language composition, education, and engineering/computer science/information technology (2017). This reflects the wide variety of course offerings available in an online setting.

Automated content analysis. While manual content analysis provides valuable, summative information from a body of literature, the amount of data analyzed is limited by the abilities of the human researchers. Therefore, a computerized content analysis could examine a large corpus of documents and do the analysis more quickly. In two identified recent research efforts, automated techniques extracted and examined distance education research. In one such study, Zawacki-Richter and Naidu (2016) produced a comprehensive report that examined 35 years of research in distance education. This study focused on published research in one journal, *Distance Education*. Text-mining techniques extracted the printed contents of titles and abstracts of 515 articles. Data analysis on 5-year increments revealed broad and developing themes. The emerging trends representing online aspects of distance education were (a) student support and (b) early phases of online education during the time period of 1995-1999 (Zawacki-Richter & Naidu, 2016). The emergence of online learning in postsecondary education was noted during the 2000-2004 period (Zawacki-Richter & Naidu, 2016). Subsequent themes included (a) online interaction patterns and (b) collaborative learning patterns for the years of 2005-2009 and (a) MOOCs, (b) OERs, and (c) interactive learning for 2010-2014. (Zawacki-Richter & Naidu, 2016).

Another study that used text mining to examine published research content involved the broader area of educational technology during the last 20 years (Natividad, 2016). For this exploratory effort, 10 highly influential journals were identified by a panel of experts in the industry. Abstracts from all contributing articles were analyzed to determine broad research areas among the nearly 10,000 articles. An emphasis on distance education was noted within the body of research (Natividad, 2016). The results "strongly emphasize learning environments that integrate technology (e.g. technology related issues, distance education, communication strategies and instructional methods, and learning experiences)" (Natividad, 2016, p. 124). Research areas were categorized within 22 clusters. One of the 22 categories created within the study is labeled as online learning. Figure 3 shows the publication trend of articles that were identified as members of the online learning factor.

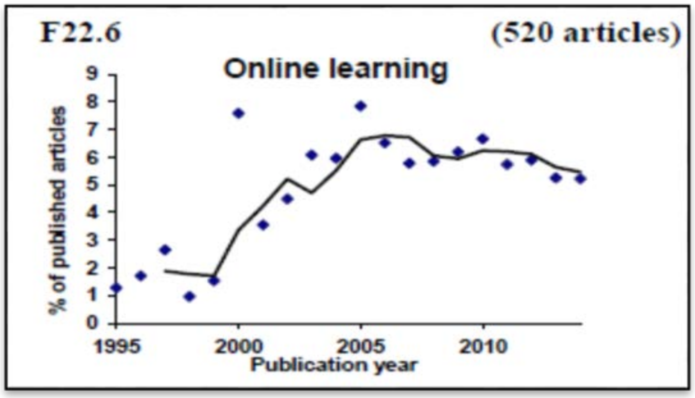

Figure 3. Online learning published articles 1995-2014 (Natividad, 2016, p. 122).

The trajectory of the published articles showed an overall increase from 1995-2005. From 2005 until 2014, the data showed a plateau or slight decline in articles within this cluster. This is surprising since online education has experienced high growth in the last decade (Franklin University, 2016). One possible explanation is that more recent articles use online access within the context of another technology focus, such as OERs, mobile, or game-based learning.

Text Mining

Using automated content analysis within the framework of meta-analysis research can provide valuable aggregated information about a larger body of research. Working with large data sets can provide information that is not available using more traditional approaches (Picciano, 2015). One way to manage large amounts of information is computer exploration to discover trends and establish meaning from the data (Feldman & Dagan, 1995). An automated analysis approach can provide an unbiased examination, as well as be a time and cost saving approach for investigation. Text mining is a viable technology that can be used to extract meaningful terms from unstructured natural language data (Hearst, 1999). Uses of text mining include classification through clustering and information extraction. Text-mining techniques have been employed to examine research subareas within the field of educational technology.

Two previously discussed studies used text mining as a part of their research process. Zawacki-Richter and Naidu (2016) used the technology when analyzing 35 years of articles taken from one distance education journal (namely – *Distance Education* edited by Som Naidu). Natividad (2016) used automated term extraction to create a latent semantic space, comprised of research areas within the broad field of educational technology. Both research efforts used automated tools to explore themes and the semantic relationship of the themes from the articles. Text mining has been employed with other educational technology topics to determine past, present, and future trends and for noting research potential gaps.

Chen, Wei, and Chen (2008) noted that content analysis is an important and versatile method for tracking and analyzing academic research, though it can be a time-consuming endeavor that relies on domain experts. They sought to develop data-mining techniques to automatically construct eLearning domain concept maps (Chen et al., 2008). Providing a pictorial depiction showing relationships between or among concepts can benefit readers, giving a more complete picture about a subject. These maps also allow for the relationships among the data to be presented in a nonlinear manner and can act as resources for educators to design and develop materials for the e-learning domain (Chen et al., 2008). The automated mechanism enabled the researchers to create content maps of e-learning research and circumvent some challenges that accompany the manual preparation of maps, most notably time and cost (Chen et al., 2008).

Another e-learning study used text mining for a different purpose. Hung (2012) sought to establish taxonomies, themes, trends, and patterns in e-learning research. The researcher reviewed 689 articles from 2000-2009 using data mining. From the data analysis, the author determined that e-learning research was in the early majority stage and that the focus of the research has since shifted from topics related to effectiveness in e-learning to topics related to teaching and learning practices (Hung, 2012). This process was extended to another aspect of e-learning—mobile technology.

An effort to examine mobile technology research was conducted by Hung and Zang (2012). The authors examined the longitudinal studies on mobile learning research published between 2003 and 2008 and determined that the most popular domains of research included (a) effectiveness, (b) evaluation, and (c) personalized systems (Hung & Zhang, 2012). They also conjectured that mobile learning as a topic within education research was in the early adopter stage, with future research focused toward strategies and framework (Hung & Zhang, 2012).

While text mining has been used in a variety of areas within educational technology research, missing from the body of published work is an effort using this technique on the subject of online learning among different journals. Using the extraction and analysis tools of a mining program on a large number of articles regarding online learning in highly regarded journals could give valuable insight into the trends of the field. This information can be used by researchers and developers to understand and produce effective online educational experiences for students.

Summary

The historical connections between online and distance learning is widely acknowledged. While this relationship is recognized, the research supports that online learning has moved beyond the affiliation and is branching out to touch other areas within educational technology. While most meta-analyses have focused on the overall topic of distance education, some studies focused solely on online education research (Allen & Seaman, 2007; Allen & Seaman, 2013; Coomey & Stephenson, 2001; C.-W. Tsai et al., 2013). Still missing is a comprehensive, automated metatrend analysis of research regarding online learning from a variety of established, highly cited journals.

Implementing automated data-mining techniques can aid with synthesizing research content, and this technology has previously been used elsewhere in educational technology. However, only one instance used automated data retrieval as a tool for conducting a metatrend analysis in online learning research, and this particular article focused on a single journal, covering 35 years (Zawacki-Richter & Naidu, 2016). Absent from the literature is an effort to identify trends and a classification model within online learning research across multiple journals using text-mining techniques.

CHAPTER 3

METHODS

The goal of this dissertation study was to identify trends in the past 20 years of research literature concerning online learning published in influential educational technology journals. Much of the meta-analyses conducted on this topic have focused on the larger subject of distance education, in which online learning could be considered a subset. Online learning typically does not refer to older delivery methods considered part of distance education, such as correspondence courses through the mail. This study aimed to identify trends in the research literature focused in online learning. The research questions addressed in this study are:

1. What are the bibliometrics of online learning articles among the 15 journals during the past 20 years (1997-2016)?
2. What are the past, present, and emerging trends of terms regarding online learning within the selected literature during the past 20 years (1997-2016)?
3. What are the thematic clusters of the articles regarding online learning in the 15 journals for the past 20 years (1997-2016)?
4. What are the time trends of the recognized thematic clusters found in the selected articles during the past 20 years (1997-2016)?

It is anticipated that by addressing these questions, an overview of online learning research emerges that proves beneficial to researchers and practitioners in the field.

This study employs an exploratory and descriptive approach to analyzing existing research literature regarding online learning. The objective is to discover past, present, and emerging trends within the body of research literature that has been published in highly regarded educational technology and online and distance education journals from 1997 to 2016. Automated text-mining techniques are utilized to analyze article abstracts, extract terms, and create categories through a statistical cluster analysis. With the use of computer programs, the text-mining process parses textual data and identifies terms within text-based documents. Text mining is similar to content analysis in that both aim to extract common trends and themes by counting words and identifying their frequency of use. Both processes also make use of computer algorithms, though text mining is "characterized by its capability of processing natural languages" (Yu, Jannasch-Pennell, & DiGangi, 2011, p. 730). Text mining adds an element of nonbiased automation to the process of analyzing text-based files. The principles of text mining strictly follow those found in other qualitative research regarding consistency and replicability, such as grounded theory and content analysis (Yu et al., 2011).

A content type analysis with a text-mining approach has been used in a variety of fields such as (a) health (Harpaz, et al., 2014; Zhu, et al., 2013), (b) international business (White, Guldiken, Hemphill, He, & Khoobdeh, 2016), (c) social media (Lazard, Scheinfeld, Bernhardt, Wilcox, & Suran, 2015; Yim & Warschauer, 2017), and, in a limited manner, for learning technologies in other specific areas such as (d) mobile technology (Hung & Zhang, 2012) and (e) e-learning (Hung, 2012). During the literature review, published work addressing the question of research trends in online learning, such as an analysis using automated text extraction techniques was missing from the corpus of publications. Presented in this chapter are discussions on the process used to obtain and analyze the data in this research effort.

Data Gathering

Data Collection

Ten educational technology journals, along with five online/distance education journals, serve as the source publications for the current study. The educational technology journals used in this study were also used in Natividad's (2016) prior study, which served as the impetus for this study. The initial 10 journals used in Natividad's (2016) research were selected by editors associated with the NTLC (see http://ntlcoalition.org/) and were identified as respected publications in the field of learning technologies. Five online/distance education journals were included to give an additional emphasis in the field of online learning. The five additional journals exhibit comparable influence in the education field. Each journal has at least a 20-year publication history. The 15 journals used in this study are:

- *British Journal of Educational Technology* (*BJET*)
- *Computers and Education*
- *Educational Technology: The Magazine for Managers of Change in Education*
- *Educational Technology Research and Development (ETR&D)*
- *Instructional Science*
- *Journal of Educational Computing Research*
- *Journal of the Learning Sciences*
- *Journal of Educational Technology & Society*
- *TechTrends*
- *Journal of Research on Technology in Education*
- *American Journal of Distance Education*
- *Distance Education*
- *Online Learning Journal*
- *Open Learning: The Journal of Open, Distance and e-Learning*
- *European Journal of Open, Distance, and E-Learning*

The data for the study was extracted from abstracts of articles with a focus in online learning. Generally excluded from consideration are (a) editorials, (b) conference reports, (c) paper discussions, (d) commentaries, (e) responses to commentaries, (f) introductions to special issues, (g) book reviews, (h) obituaries, (i) errata, and so forth. However, some of these excluded categories of content have the length or features similar to an article and could possibly have been retained for inclusion in the study.

The data from this research was assembled from a variety of public access sources. Information was accessed from a prior study in educational technology using 10 of the journals for the years 1997-2014 (Natividad, 2016). Additional article data from these journals for the two most current years, 2015 and 2016, and the other five journals for the entire 20-year period were collected through electronic databases and e-journals, with one exception. One journal from the original 10, *Educational Technology: The Magazine for Managers of Change in Education* is available only in print, and issues from 2015 and 2016 were scanned to form pdf files. Data from every obtained article was entered into a Microsoft Excel (Excel) spreadsheet with each row containing data for one article. Data for each journal remained in a separate file throughout the cleaning.

Data Cleanup

Using functions in Excel, the data was reviewed, and any article which did not meet the initial inclusion criteria was eliminated from consideration. In the Excel file, each article's data was represented by a row. The initial process for data clean-up involved:

1. The page length of each article was determined by calculating the difference of the starting/ending page numbers and adding one. Calculating the page length helps in determining which articles may be editorials, introductions, conference advertisements, errata, and so forth.

2. Each of the entries was checked for suitability as a journal article. Rows that did not appear to be journal articles, such as editorials, book reviews, and so forth, were flagged for deletion.
3. The citation and abstract for each article was verified. Some databases contain unique abstracts for articles. The protocol for the abstracts in this study was as follows:
    a. The abstract from the article or the journal Website was used.
    b. If an abstract was not available via the article itself or the journal's Website, then an abstract from an academic database, such as ERIC or Ebscohost, was used.
    c. If no abstract was found, then the article was re-evaluated for inclusion. If the article was to be included, then an abstract was written using text within the article, usually from the introductory or concluding paragraphs.

Once the initial processing was complete, articles that did not meet the criteria, such as reviews and editorials, were removed from the data set. Each article that appeared to be outside the parameters of a journal article was examined to determine whether it contained information representative of an article. Example of deleted entries include book or product reviews, editorials, advertisements, errata, and so forth. If the article was judged as an appropriate article, it remained in the data file. Otherwise, that article was deleted. This process netted 12,356 articles from the 15 journals to be considered for this study.

The final step in the data clean-up process involved finalizing the journal data in the file by sorting the remaining columns to identify missing data (missing author, page number, etc.). Any missing information was manually corrected by finding the article and modifying the information contained in the file. Once final changes were made to the file, each article was given a unit identification code. Another column for identification was added titled *Article ID*. Each article was assigned an alphanumeric code. The first digits are letters corresponding to the sponsoring journal. The last four digits contain numbers identifying that article within the journal. Once this process was complete, the data from the individual journals were combined into one Excel spreadsheet file. The spreadsheet contained 12,357 rows, one for each article plus one for column labels.

Data Selection

The total number of articles from the 15 journals considered for this study was 12,356. Once all the articles from the 15 journals had been through the data clean-up process, they were finalized in the spreadsheet and the ones used in this study were selected. The overarching criterion for article inclusion in this study was that a focus of online learning must be evident. A variety of search strings was used to draw out articles that relate to this topic. Search strings used in the titles of the articles were *online, on-line, on line, distance, internet, e-learning, open, MOOC, virtual,* and *web.* Search strings used in the abstracts were *online, on-line, on line, distance education, distance learning, and MOOC.* Any entry row not containing any of the search text in the title or abstract cell of the spreadsheet was deleted from the Excel file. The number of articles remaining totaled 5,151 and this group was further considered for this study. Table 3 shows the number of articles during each step of the selection process.

Table 3

Article Counts from the 15 Journals

Journal Name	Total Articles (1997-2016)	After Search	Selected Articles
British Journal of Educational Technology	1472	568	556
Computers and Education	2666	1032	1001
Educational Technology Magazine	986	293	278
Education Technology Research and Development	725	201	183
Instructional Science	570	88	78
Journal of Educational Computing Research	865	294	274
Journal of Educational Technology & Society	1467	602	581
Journal of the Learning Sciences	272	21	17
Journal of Research on Technology in Education	456	114	105
Tech Trends	921	310	296
American Journal of Distance Education	298	267	254
Distance Education	406	354	315
European Journal of Open, Distance, and E-Learning	338	252	229
OLC Online Learning Journal	532	441	440
Open Learning	382	314	214
Totals	12,356	5,151	4,821

All of the remaining 5,151 articles were vetted for suitability for inclusion in this study of online learning. For each of the articles, the title, abstract, or body of the article was evaluated by the researcher to ascertain appropriate content regarding online learning. Even though a title or an abstract contains an element from a search string, the article's relation to online learning is not certain. Of the articles retrieved from the searches, 330 did not contain a focus in online learning and were removed from consideration in this study. The number of articles remaining was 4,821 and this group was used for the current study (see Table 3).

Reasons why articles retrieved using the search string were omitted from the study varied. However, the commonality among them was the absence of evidence pertaining to online learning. Some omitted distance education articles were from earlier in the time period and used other forms of instructional delivery such as VHS tapes or paper correspondence. Some of the distance and open education articles excluded were based in remote geographical areas which, at the time of the article writing, did not possess the technological infrastructure to support Internet or networking services. If an article abstract mentioned a digital, electronic, or computerized component but did not clearly indicate an online component, the full article was checked. The articles not selected included those running on a standalone device, such as an older desktop computer with no networking capabilities. In such an instance, the learner received the software possibly by computer disks through the mail or factory installations, but not through any type of network. If the online component of an article was deemed uncertain, the article was excluded.

Once the 4,821 articles were vetted, the last step was to create a new identifier for each one. In order to better organize and track the data during analysis, a new alpha numeric identification code was assigned to each article. The first characters representing the journal and the subsequent digits represented the article in order of appearance in the journal. For example, article AJDE0001 would represent an online learning article published in the *American Journal of Distance Education* during 1997, the first year of the 20-year period for this study. The article AJDE0254 represents and online learning article from the journal published in 2016. Using a coding system enabled quick identification of articles if needed during analysis.

Data Analysis

This section describes the methods used to properly address the research questions. Each question is unique and requires different processes to address it properly. The research questions and subsequent methods were considered in the order given.

Bibliometric Analysis

Research Question 1 states: What are the bibliometrics for the selected articles? Bibliometric analysis provides a summary of a body of literature by measuring selected indicators (Thelwall, 2008). Bibliometric measures allow researchers to generate quantitative information from large amounts of historical data (Hung, 2012). The purpose of the bibliometric analysis in this study was to give a basic overview of online education article publication dates and corresponding journals, providing a set-up for the research analysis examining trends in the literature that followed. Bibliometric measures for this study included prolific journal rankings of related articles with percentages and yearly publication trends of the online learning research within the journals. Table 4 contains the number of articles included in the study from each journal and the corresponding percentage as related to the entire collection of articles across the 15 journals. The online learning articles comprise slightly over 39% of the total articles from the 15 journals.

Bibliometric information about the online learning articles from each journal as it pertains to articles used in this study is displayed in Table 5. The journals are listed in descending order based on the quantity of articles contributing to the study. The most prolific journal is *Computers and Education,* which houses approximately one fifth of the selected articles. The least prolific journal is *Instructional Science*, which contributed less than 1% of the article data.

Table 4

Journal Article Tally and Percentage

Journal Name	Total Articles (1997-2016)	Online Learning Articles	% of Online Learning Articles
Computers and Education	2666	1001	37.55%
Journal of Educational Technology & Society	1467	581	39.60%
British Journal of Educational Technology	1472	556	37.77%
OLC Online Learning Journal	532	440	82.71%
Distance Education	406	315	77.59%
Tech Trends	921	296	32.14%
Educational Technology Magazine	986	278	28.19%
Journal of Educational Computing Research	865	274	31.68%
American Journal of Distance Education	298	254	85.23%
European Journal of Open, Distance, and E-Learning	338	229	67.75%
Open Learning	382	214	56.02%
Education Technology Research and Development	725	183	25.24%
Journal of Research on Technology in Education	456	105	23.03%
Instructional Science	570	78	13.68%
Journal of the Learning Sciences	272	17	6.25%
Totals	12,356	4,821	39.02%

Table 5

Online Learning Article Tally and Percentage of Online Learning Articles

Journal Name	Online Learning Articles	% of Total Selected Articles
Computers and Education	1001	20.76%
Journal of Educational Technology & Society	581	12.05%
British Journal of Educational Technology	556	11.53%
OLC Online Learning Journal	440	9.13%
Distance Education	315	6.53%
Tech Trends	296	6.14%
Educational Technology Magazine	278	5.77%
Journal of Educational Computing Research	274	5.68%
American Journal of Distance Education	254	5.27%
European Journal of Open, Distance, and E-Learning	229	4.75%
Open Learning	214	4.44%
Education Technology Research and Development	183	3.80%
Journal of Research on Technology in Education	105	2.18%
Instructional Science	78	1.62%
Journal of the Learning Sciences	17	0.35%
Totals	4,821	100.00%

The third and final bibliometric analysis performed is based on yearly publications. The number of online learning articles published each year along with the corresponding percentage for that year is shown in Table 6.

Table 6

Total and Percentage of Online Learning Articles Published Each Year

Year	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
# of Articles	82	88	115	194	200	173	185	216	212	207
% of All Articles	22.7%	23.6%	27.4%	40.3%	44.6%	33.7%	40.7%	43.2%	39.0%	40.0%

Year	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
# of Articles	261	293	335	336	299	359	320	324	338	284
% of All Articles	42.2%	40.1%	44.9%	43.0%	39.7%	43.0%	36.4%	40.9%	39.5%	37.6%

The year with the lowest measure is the first year of the time period (1997), with 82 articles published that year, comprising 22.7% of the articles in the 15 journals. Although year 2012 had the most online learning articles published with 359, the highest percentage of online learning articles published can be found in the year 2009 with 44.9%. It is noted that for the all but the first three years, the percentage of online articles varied but suggest a stabilization in publication ranging between 33.7% and 44.9%. Information from this bibliometric data is used to establish a foundation for the remaining research questions.

Term Trends Analysis

Research Question 2 states: What are the past, present, and emerging trends in the terms used in the research? Text-mining techniques are used to extract the terms for analysis. The method in this section generally follows the steps in Denning, Fisher, and Higgens (2011) for extracting terms. The extracted article abstracts provide data source for the terms. The abstracts are a good representation for the content of the full articles. It is expected that the abstracts are carefully written by the author to reflect the general content and findings reflected in the full article. Also, abstracts are shorter in length than a full article, which not only eliminates data noise but also allows the analysis to be more focused and efficient. RapidMiner Studio (2017), an open source data science platform, was used to extract the terms and provide the automated analysis and has been utilized in other text-mining published projects (Denning et al., 2011; Martens, Poeppelbuss, & Teuteberg, 2011).

In order to quantitatively analyze text-based documents, vectors were created. The procedure used to create the vectors is shown in Figure 4.

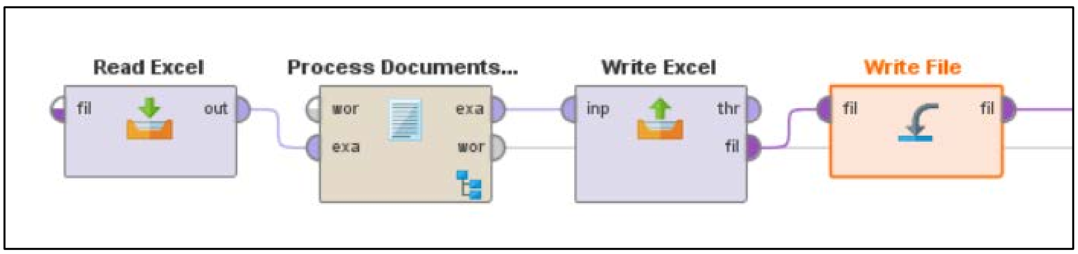

Figure 4. Method for extracting terms in RapidMiner (2017).

To create word vectors, the master Excel file was read by an operator in RapidMiner (2017). The file is a spreadsheet containing 4,822 rows, one for each online learning article plus an additional row for column headings. The data was input to a process document operator, which generated initial word vectors. Inside this operator, the data goes through eight steps to clean the data so that meaningful terms can be counted. Figure 5 displays these eight steps. After the text data was processed, the output was transformed into a Microsoft Excel compatible format and then written to a file.

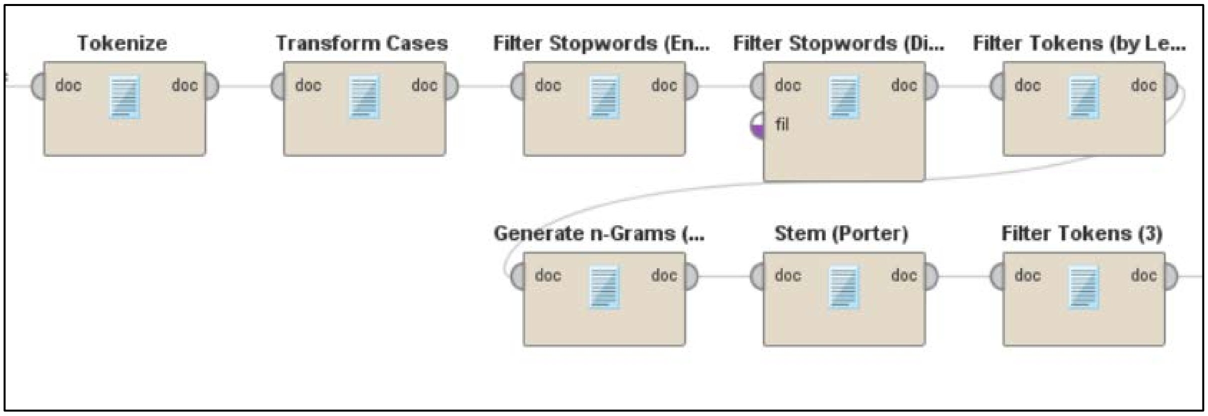

Figure 5. The eight operators used to process abstract data (RapidMiner, 2017).

An 8-step process (see Figure 5) was used to create the word vectors used in the analysis. The stages are (a) tokenize, (b) transform cases, (c) filter stop words from default list, (d) filer stop from focused list, (e) filter short words, (f) generate n-grams, (g) stem n-grams, and (h) filter short n-grams. An explanation for each step is detailed:

1. Tokenize: The text data is read and converted into unique digital tokens.

2. Transform cases: All letters are converted into lower case so the tokens with the same letters will be considered the same word. For example, the words *teach* and *Teach* are considered the same because the capital T in the second word is converted to lower case.

3. Filter stop words (English) – Common words contained in the RapidMiner (2017) stop word list, such as pronouns, are deleted from the data set. These words do not add meaningful context to the data and can potentially water down the data set.

4. Filter stop words (Dictionary) – These are additional words deemed as needed stop words for this study. The default list from RapidMiner (2017) was not located, so to make certain all necessary stop words were incorporated, an additional list was compiled to ascertain the remaining terms were meaningful to the study. The additional stop words were adopted from a widely used English stop word list (Lextek International, 2002) and a previous educational

technology document analysis study (Natividad, 2016). A list of the stop words can be found in Appendix C.

5. Filter tokens (by length) – Tokens with a length of less than three characters are deleted from the data.

6. Generate n-grams – In order to capture two- or three-word phrases, called n-grams, within the texts, the data is run through an n-gram operator. The maximum n-gram length of the operator is three tokens. Generating n-grams enables the tracking of terms with multiple words such as *discussion forum* and *chat room*.

7. Stem (Porter) – This operator stems English words using the Porter stem algorithm (Willett, 2006), which replaces word suffixes to permit different tenses of the same word to be recognized as the same term. For example, the words *discuss*, *discussing*, and *discussed* are all recognized as the same term.

8. Filter tokens (3) – The stemmed n-grams are filtered one last time in order to eliminate any strings with a character count of less than three. All cleaned tokens that remain consist of three or more characters.

The output for the 8-step process document is a spreadsheet with four columns. A snapshot of the spreadsheet is displayed showing the first five rows (see Figure 6).

Word	Attribut...	Total O...	Docum...
abil	abil	22	21
abl	abl	14	13
absenc	absenc	4	4
abstract	abstract	5	3
academ	academ	90	59

Figure 6. Output example of process document operator from RapidMiner (2017).

The first column (Word) is the term and the second column is titled attribute. The third column (Total O…) indicates the number of times, or total occurrences, a specific term is used in the text data body. The fourth column (Docum…) indicates the total number of documents that contain the term.

The RapidMiner (2017) software program was used to analyze the entire data set and also within the four 5-year periods (1997-2001, 2002-2006, 2007-2011, and 2012-2016). Using this process allowed a more robust analysis of the trending terms as related to a specific period of time and possibly lead to a better overall analysis of past, present, and emerging trends of terms.

Creation of word clouds for each of the 5-year segments provided a visual graphic depiction of the trends in the terms used in the literature. Word clouds provide a unique way to summarize the content of text documents (Cui, et al., 2010). Word clouds are formed by the words contained in the textual data. The size of a word in the cloud is proportional to the number of times the word is used throughout the data set. However, to avoid confusion when looking at the word clouds, the actual words of the abstracts are used as opposed to stemmed n-grams. The process operator in RapidMiner (2017) for creating word cloud lists is similar to that of the process for creating n-grams, minus the stem and n-gram pieces so that whole words are used, as shown in Figure 7. One filter token operator is needed at the end of the process to allow only words that comprise three or more characters to be included in the analysis.

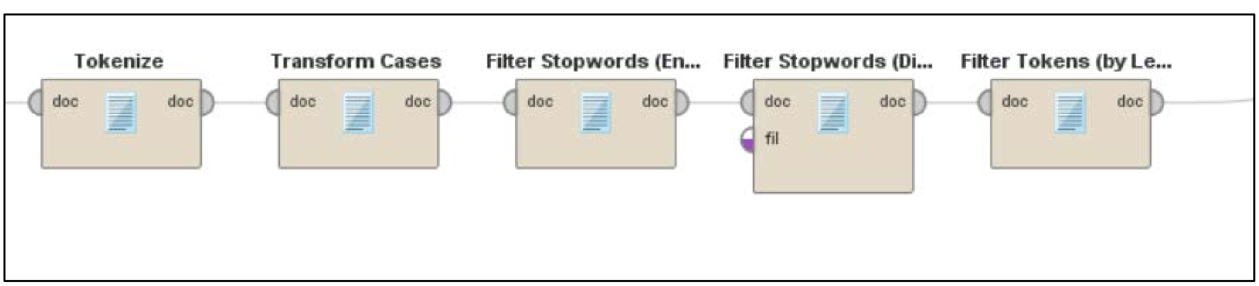

Figure 7. Process document operation in RapidMiner (2017) for whole words.

Using the word count output from RapidMiner (2017), the word clouds were created with the online program, Wordle (Feinberg, 2014). The 100 most frequently used words in the article abstracts comprise the clouds. A word cloud was created for each of the four 5-year periods. To conclude the examination of trending terms, a word cloud covering the entire 20-year span was also formed.

Cluster Analysis

Research Question 3 states: What are the thematic clusters of the articles? Cluster analysis is a popular technique in text-mining studies (Delen & Crossland, 2008) and has been used in previous literature analysis (Denning et al., 2011; Durmuşoğlu, A., 2016; White et al., 2016). Cluster analysis is one method of LSA (Evangelopoulos et al., 2012) and is used to create categorical clusters from textual data. LSA is a "technique for comparing texts using a vector-based representation that is learned from a corpus" (Wiemer-Hastings, Wiemer-Hastings, & Graesser, 2004, p. 1). Clustering divides the collection of articles into mutually exclusive groups based on themes detected within the text-based document (Chakraborty et al., 2014).

It should be noted that in the previous study of 10 educational technology journals (Natividad, 2016), LSA was performed on the data to create the categories of data. However, the algorithm used was a type of factor analysis, not a cluster analysis. While evidence of the viability and history of clustering has been presented, factor analysis was a possible option. The choice of clustering over factoring was due primarily to available resources. The RapidMiner Studio (2017) platform is open source, so access to the program is available at no cost. This researcher used the k-means clustering operators within RapidMiner (2017) for prior presented work and was confident in the dependability and consistency of this analysis. To perform the exact same analysis as used in Natividad's (2016) research, a completely different program would be required. The lead researcher for the current research could not access the program without a significant cost (several thousands of dollars) or having the data run through a third party. By using Rapid Miner (2017), the researcher was able to retain full control over the data—before, during, and after running it through the automated processes within the data-mining program.

A k-means cluster analysis produces a single level of groupings. This statistical method is one of the clustering techniques that can be used to bring understanding to unstructured data (Yang & Chen, 2002; Pushplata, 2012). The k-means method for developing clusters from text data has been used in prior published automated content analysis (Denning et al., 2011; Martens et al., 2011). The overall methodology utilized in performing the cluster analysis is adapted from the five methodological considerations outlined in Evangelopoulus et al. (2012) article, "Latent Semantic Analysis: Five Methodological Recommendations" which include (a) term filtering, (b) term weighing, (c) dimensionality reduction, (d) threshold selection, and (e) post-LSA quantitative analysis.

Cluster creation. The first consideration for cluster creation is term filtering, which is completed through the prior effort in this study during the term extraction used to address the second research question. The second consideration is *term weighting,* which calculates the importance of a term to a document. For this step, the *term frequency-inverse document frequency value* (tf-idf) was calculated for each term within each article. Instead of a total of term occurrences in the previous procedure (with the creation of the word vector Excel file), the tf-idf algorithm computes values for each term from textual data "through an inverse proportion of the frequency of the word in a particular document to the percentage of documents the word appears in" (Ramos, 2003, p. 1).

The third consideration is *dimensionality reduction*, which reduces data so that the original value is maintained but takes up less digital storage space. The RapidMiner (2017) default parameters in the cluster program were used for this step. The fourth consideration is *threshold selection*. This step is a set-up to determine the proper number of clusters, *k*, to form from the k-means clustering process. A plot of group sum of squares by the number of clusters provides a graph that helps determine a suitable number of clusters. Similar to a scree test in factor analysis, an elbow bend in the graph denotes an appropriate number of clusters (Kabakoff, 2017). To determine the appropriate number of clusters, code created under the statistics program *R* (R Foundation, 2017) was executed. Code for this process was adapted from an example ("Determining the optimal number of clusters", n.d.) and is displayed in Figure 8.

```
> set.seed(123)
> OnlineLearn <- read.csv(file.choose())
> data <- OnlineLearn
> k.max<-30
> wss <- sapply(1:k.max, function(k){kmeans(data,k, iter.max=30)$tot.withinss})
> plot(1:k.max, wss,
+        type="b", pch = 19, frame = FALSE,
+        xlab="Number of clusters K",
+        ylab="Total within-clusters sum of squares")
```

Figure 8. Code in *R* used to create sum of squares graph (R_Foundation, 2017).

The graph for the sum of squares for 4,821 article abstract clusters is shown in Figure 9. In the graph shown in Figure 9, more than one steep drop or elbow is evident in the graph. For this study, the number of clusters, the k value, was set at 18 as indicated by the arrow in the diagram. There is another significant decline preceding the number 10. A test k-means cluster run was completed setting the k value to 10. However, some of the clusters combined in a way that prompted a vaguer description of the data. When the cluster number was set to greater than 18, the topical description did not meaningfully improve. The larger clusters appearing when the k value was 18, did not break into smaller more descriptive topical groupings. The number, 18, was determined to be an appropriate number of categories to effectively express the article abstract data.

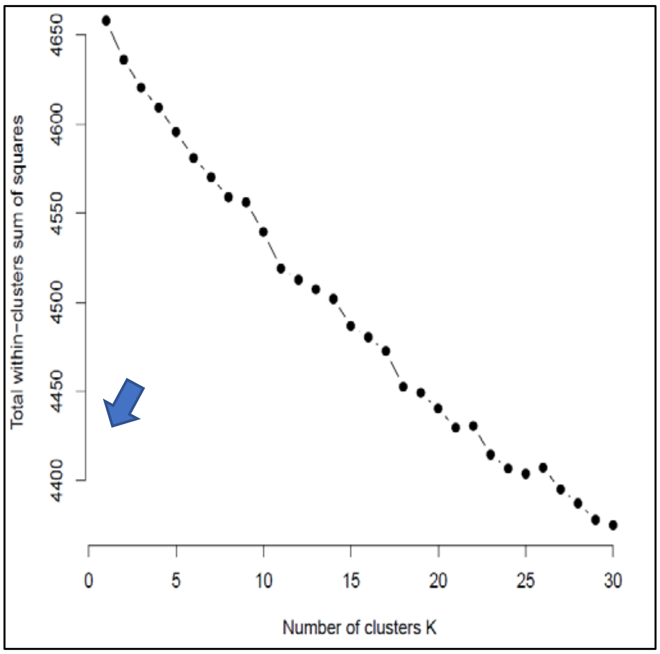

Figure 9. Graph of sum of squares vs. number of clusters.

The cluster analysis was run within RapidMiner Studio (2017). The article abstract data was processed through five main operators to create the 18 clusters. A screenshot from RapidMiner (2017), showing the cluster analysis phases, or steps, is displayed in Figure 10.

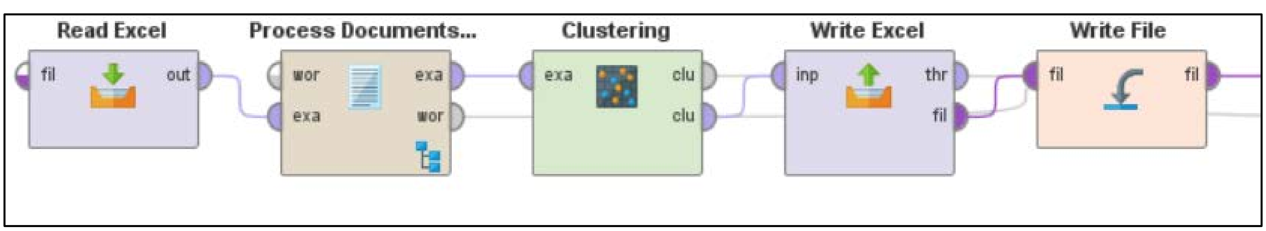

Figure 10. The 6 steps of cluster analysis.

The first step in the process as shown in Figure 10 is to read the Excel data file that was prepared after the data clean-up phase. The steps inside the process operator are similar to the previously described process document operator. In this case, the terms are stemmed but n-grams are not necessary to produce data needed for clustering. The output from the process documents is a table with tf-idf values for each term in each article. An example for this file is shown in Figure 11.

	AY	AZ	BA	BB	BC	BD	BE
1	**activ**	**actor**	**actual**	**actuat**	**acut**	**adapt**	**add**
12	0.06134	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
13	0.05119	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
14	0.11242	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
15	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
16	0.00000	0.00000	0.00000	0.00000	0.00000	0.09319	0.00000
17	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
18	0.12711	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
19	0.05157	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
20	0.00000	0.00000	0.00000	0.00000	0.00000	0.06925	0.00000
21	0.19701	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
22	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
23	0.00000	0.00000	0.00000	0.00000	0.00000	0.12599	0.00000
24	0.00000	0.00000	0.11140	0.00000	0.00000	0.00000	0.00000

Figure 11. Screen shot of tf-idf values.

The far left column displays the article row number for the original Excel file containing the 4,821 articles. In the example shown, articles in rows 12, 13, 15, 18 and 21 all contain words with the stem *active*, with the article in row 21 having the strongest representation of that stem. The article housed in row 24 contains at least one word with the stem *actual*. Articles represented in rows 16, 20, and 23 have words that are comprised of the stem *adapt.* For articles in rows 12-14, no words based on stems *actor* and *add* appear.

The tf-idf table acts as input for the clustering operator. For the k-means clustering process, the *k* value is set to 18, and the initial seed and max iterations are set to the same values used in within-sum-of squares code. One output for the cluster operation is an Excel file with the rows representing the articles and the columns representing the stemmed terms (see Figure 12).

	AY	AZ	BA	BB	BC	BD	BE	GHP	GHQ
1	**activ**	**actor**	**actual**	**actuat**	**acut**	**adapt**	**add**	**id**	**cluster**
12	0.06134	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	11.00000	cluster_11
13	0.05119	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	12.00000	cluster_17
14	0.11242	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	13.00000	cluster_6
15	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	14.00000	cluster_10
16	0.00000	0.00000	0.00000	0.00000	0.00000	0.09319	0.00000	15.00000	cluster_8
17	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	16.00000	cluster_5
18	0.12711	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	17.00000	cluster_6
19	0.05157	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	18.00000	cluster_11
20	0.00000	0.00000	0.00000	0.00000	0.00000	0.06925	0.00000	19.00000	cluster_6
21	0.19701	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	20.00000	cluster_3
22	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	21.00000	cluster_4
23	0.00000	0.00000	0.00000	0.00000	0.00000	0.12599	0.00000	22.00000	cluster_8
24	0.00000	0.00000	0.11140	0.00000	0.00000	0.00000	0.00000	23.00000	cluster_17

Figure 12. Screen shot of Excel file with tf-idf values and cluster assignments.

The first seven columns are identical to those shown in Figure 11 and serve as input to the cluster operators. The calculated cluster dimensions are added into the last two columns of the Excel spreadsheet. The graphic in Figure 12 intentionally hides the columns between BE and GHP in order to fit the example on the page. The second-to-last column is the program ID for the article. The numbers range from 1 to 4,821 and are assigned in the same order as the input data file. The last column indicates the cluster number assigned by the program. Each article is assigned to a cluster. A screenshot of the cluster program output is given in Figure 13.

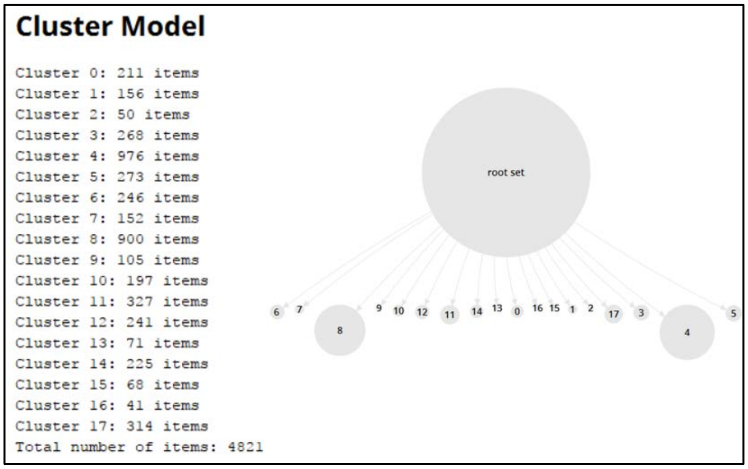

Figure 13. Output cluster model from RapidMiner (2017).

A second type of output from the clustering program is a centroid table. A cluster is created based the proximity of terms within articles to the center, or centroid. A centroid table contains values for each term in the member articles based on their tf-idf weight for that cluster. Tables for the clusters are provided in Chapter 4, along with a discussion for each cluster.

Cluster labeling. The final step is labeling the clusters. Each cluster is assigned a meaningful identification code and corresponding label. While the labeling of clusters is subjective by nature, steps were taken to appropriately designate each cluster to represent the articles assigned to it. To name the clusters, the researcher examined multiple types of data within each cluster. This process included looking at the top terms within each cluster, titles of the cluster member articles, and possibly, reading the abstracts of each associated article. Cluster labels were coded sequentially in order of the average publication year of member articles.

Cluster Trend Analysis

Research Question 4 states: What are the time trends of the clusters? This information is calculated by weighting the clusters for each year according to the number of articles written that year. The trends for publication are viewed from multiple perspectives. Descriptive statistics were used to calculate the yearly mean, median, and mode for each cluster. Also, two graphs for each cluster were created to further illustrate trends in the publications. The first graph depicts the number of articles published each year in the cluster. The second graph denotes the percentage of online learning articles published each year for the clusters. The two graphs are shown in Chapter 4 along with a discussion of each cluster.

Summary

The initial data field consisted of 12,356 articles from 15 journals for a 20-year period (1997-2016). Out of that body, 4,821 articles were identified as having a direct relationship with the topic of online learning. Various analyses were performed using the 4,821 articles and their abstracts in an effort to address the four research questions.

A bibliometric analysis was performed on the body of documents to provide foundational information about the articles to address Research Question 1. Specifics include a tally of articles in each journal along with the percentage of online learning articles compared with the entire corpus. The total number and percentage of articles published each year was also investigated.

To find trends in terms that address the Research Question 2, abstracts from all 4,821 articles were examined. A text-mining program in RapidMiner (2017) was employed to extract and organize the data. Each operator within the program and its role was discussed. Examples of the n-gram output was shown and explained. Five different runs were conducted to discover trends within the 20-year period, four of the runs were over 5-year periods and the final one included all 20 years. The breakdown in years of the runs follows: 1997-2001, 2002-2006, 2007-2011, 2012-2016, and 1997-2016.

Research Question 3 involves categorizing the articles through clustering. A k-means clustering program in RapidMiner (2017) was used to assign each article to one of 18 unique clusters. The operators used in this process were explained. Output from the program was presented and described. The output calculations from the cluster run were used to address the Research Question 4, concerning time trends among the clusters.

Results from the analyses can be found in Chapter 4. The outcomes as they pertain to the four research questions are discussed. A variety of tables and figures presented in Chapter 4 provide multiple representations of the results.

CHAPTER 4

RESULTS

In this chapter, results from the data analysis are presented to begin establishing longitudinal trends in online learning research data. Outcomes from the analysis are depicted and discussed. The results from this study produced several interesting outcomes related to research in online learning. Presenting all possible inferences, though informative, would produce and abundance of information which stretches or exceeds the bounds for this research. As such, the findings presented in this chapter were intentionally narrowed. Decisions on what to present and its format were gauged on how effectively the results and corresponding presentation format addressed the research questions. The four research questions regarding online learning research (1997-2016) directing this study involve analyzing the (a) bibliometrics, (b) trends in terms, (c) thematic clusters, and (d) time trends of the clusters. The results from the bibliometric analysis is given first followed by the outcomes from the trending term analysis. Lastly, results from the cluster analysis and cluster time-trend investigation are presented.

Bibliometrics

The results in this section address the first research question concerning the bibliometrics of the 4,821 articles that met the selection criteria for this study: What are the bibliometrics of online learning articles among the 15 journals during the past 20 years (1997-2016)?

Overview of the 15 Journals

To give suitable background and meaning about the online learning contributions from each of the journals, it is worth mentioning the general publication numbers for the journals. The graph in Figure 14, displays the total number of articles in each journal (the longer bar) and the total number of articles regarding online learning (the shorter bar). With *Computers and Education* having the most issues and subsequently the most articles at 2,666 published, it follows that the journal would contribute more articles (1,001). Some journals have a fewer articles and issues, such as *Journal of the Learning Sciences* (272 articles) and American *Journal of Distance Education* (298 articles). As such, these journals contribute a smaller number that some of the other journals. However, because the mission of the *American Journal of Distance Education* closely aligns with the focus of the study, it contributes considerable more articles (254) than *Journal of the Learning Sciences* (17). The graph is meant to give context to the overall contribution of online learning articles percentage of each journal.

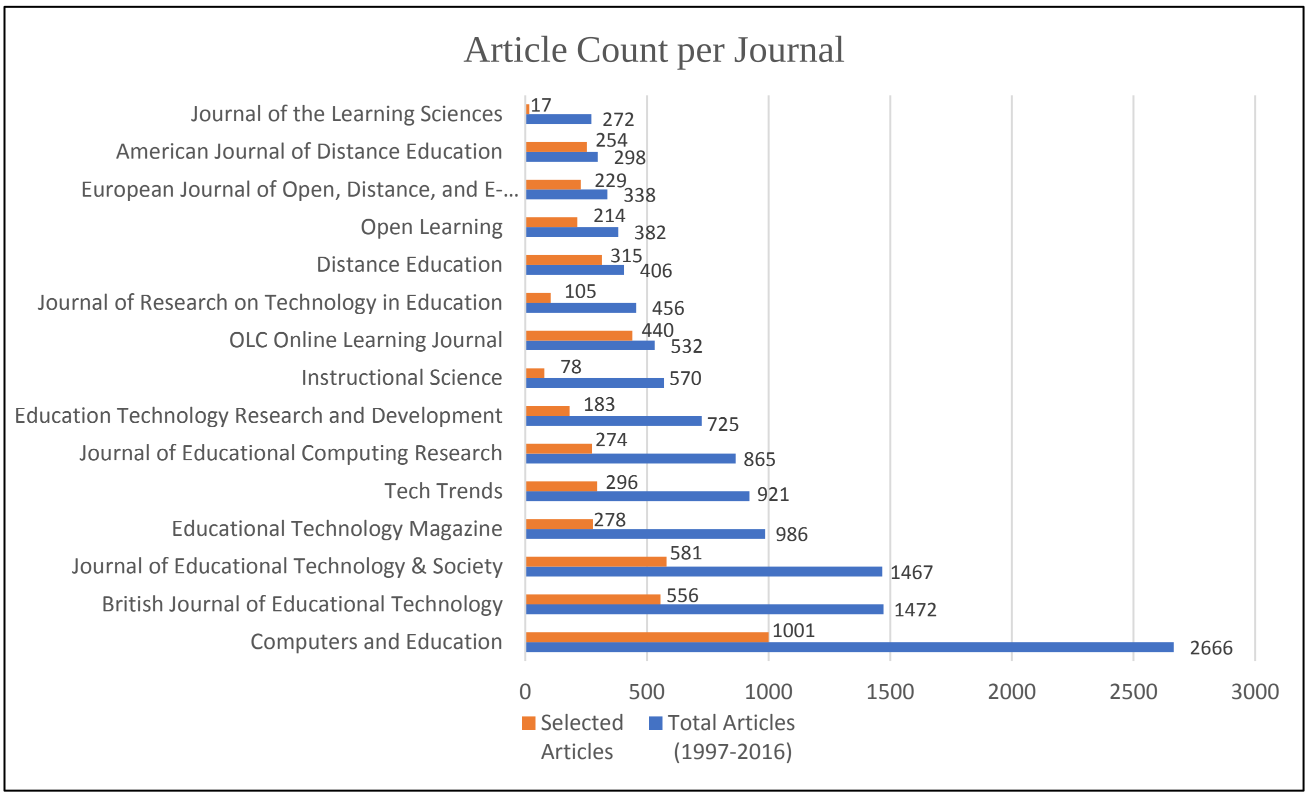

Figure 14. Bar Graph showing number of articles from each journal

The graphic shown in Figure 15 provides an overview of contribution by denoting each journal selected for the study and the corresponding percentage of articles contributed. Of the 15 journals, 10 focus on education and/or technology, and five maintain a more specific focus in online or distance learning.

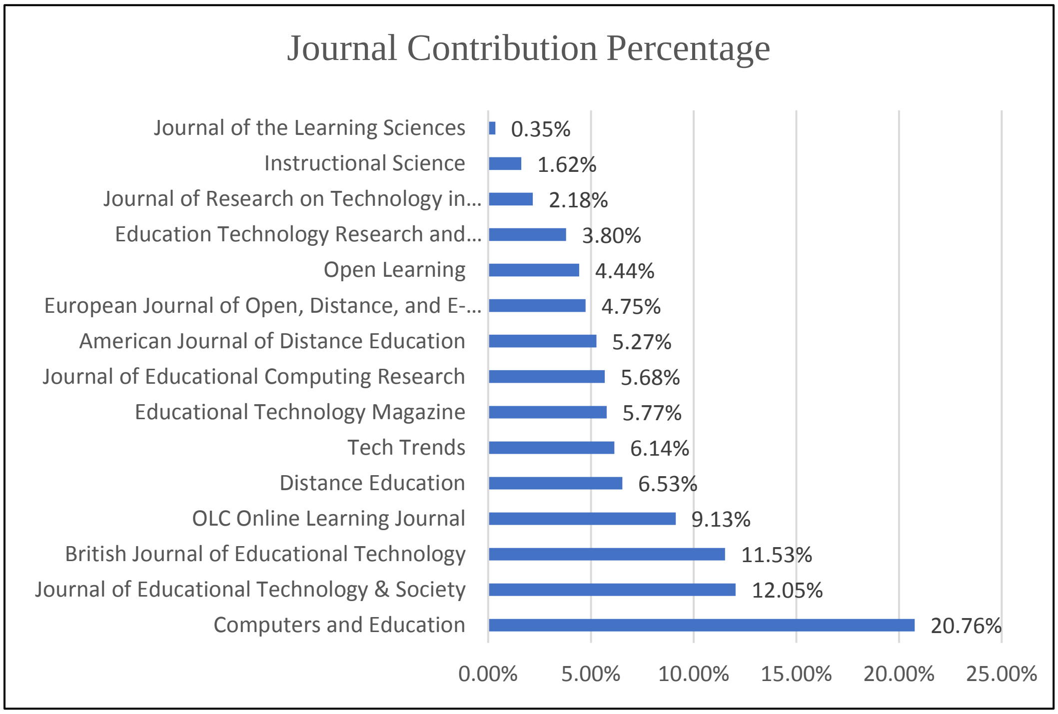

Figure 15. Graph depicting journal contribution percentage.

Each of the 15 journals is sequentially listed from least-to-greatest percentage contribution. Of the five journals with the greatest number of contributions, three are from the original educational technology journal list and two are from the more specialized online learning/distance education list. This is not unexpected as the educational technology journals generally have substantially more published articles. The journal with the highest contribution, *Computers and Education*, contains 2,666 articles with less than half (1,001) related to online learning. The five journals specializing in distance and/or online learning have fewer published articles. The journal from that group contributing the most articles to the study is *OLC Online Learning Journal*, which published 532 recognized articles during the 20-year period and over

82%, 440, of those qualified for this study. Therefore, this indicates that online learning maintains a considerable role in the field of educational technology, as well as, that research published in educational technology journals influences the course of online learning research.

Historical Bibliometrics

The trend of the online learning articles compared with articles written on other educational technology topics is shown in Figure 16. The upper portion of the graph represents online learning articles published and the lower potion denotes the number of articles written on the other topics. This provides perspective on how the research regarding online learning has grown compared to the rest of the educational technology field. The online learning articles generally show a steeper increase in earlier years of publication when compared to the aggregate of the other topics. However, for most of the time the publication trends tend to mirror.

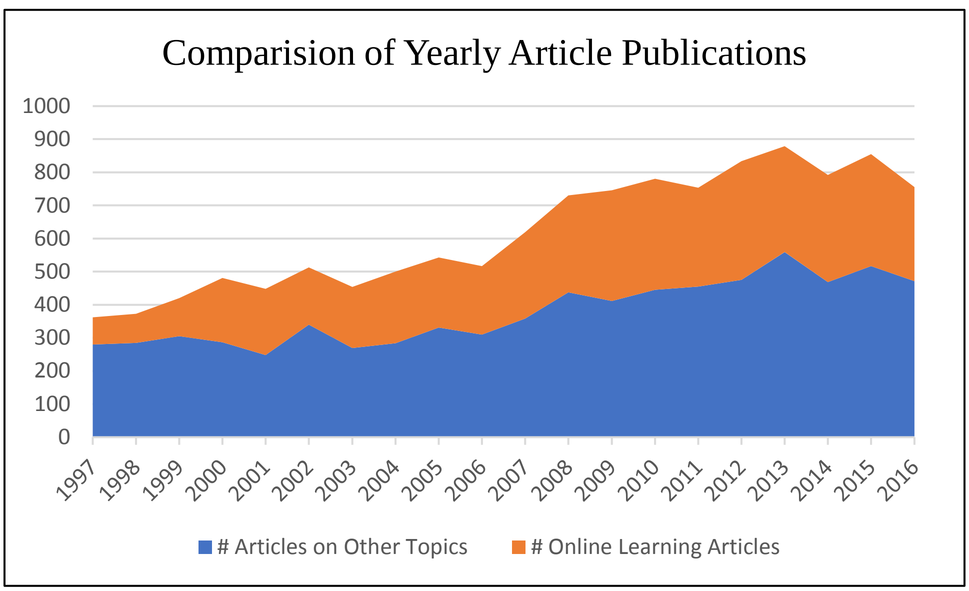

Figure 16. Graph comparing online learning articles with the total number of articles from the 15 journals.

A more focused analysis of the overall yearly publications of online learning articles is provided in Figure 17. The graph depicts how many articles related to online learning were published from the 15 journals each year. The solid line in the graph connects the data points and the dotted curve follows the moving average of the data points suggesting trends within the data.

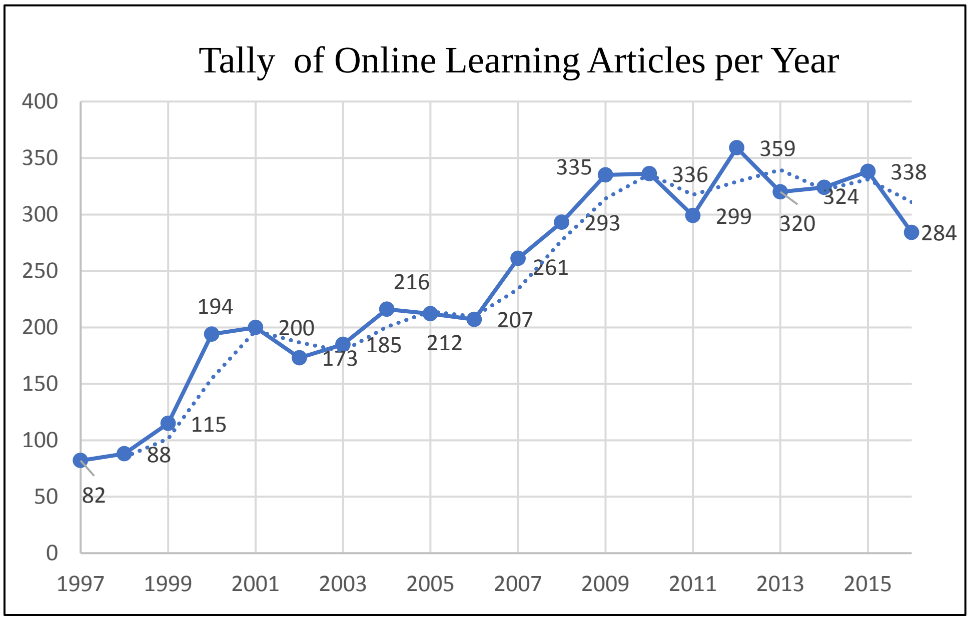

Figure 17. Graph depicting number of online learning articles published each year.

The difference between the lowest publication year, 1997, at 82 articles to the highest publication year, 2009, at 359 articles represents a 337.8% increase. The graph shows a general increase in publications through 2009 and then the graph seems to level or perhaps demonstrates a general decline. This could indicate a decline in publishing, or, as online learning becomes more commonplace, this graph could indicate an assumption of an online presence without mentioning it in the article abstract.

When compared to the percentage of articles published in the journals a similar trend develops. Presented in Figure 18 is the yearly percentage of online learning articles published during each year in the 15 journals. In this graph, the percentages are calculated by dividing the number of online learning articles published in a given year by the total number of articles published the same year in the 15 journals.

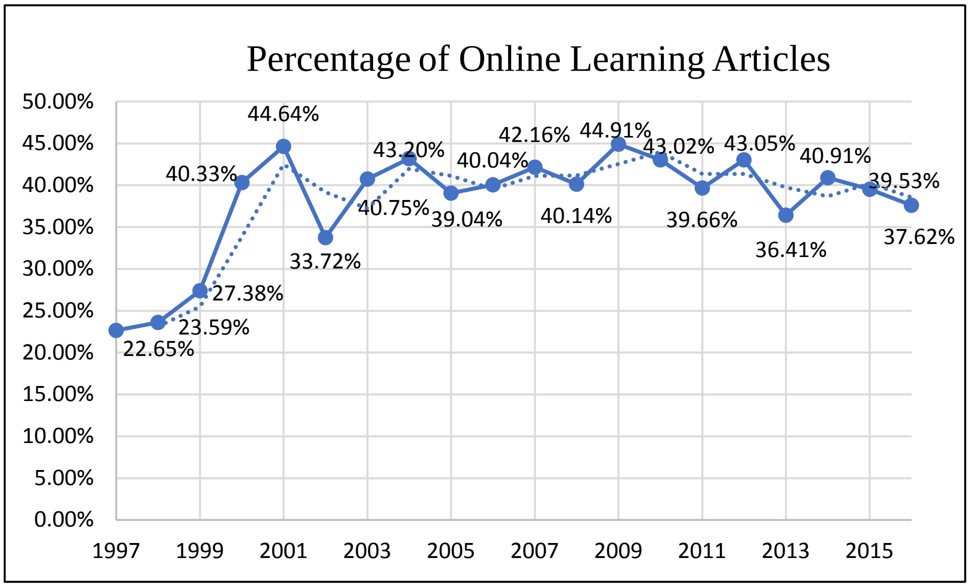

Figure 18. Percentage of articles related to online learning compared with total articles published each year in the 15 journals.

The lowest percentage can be found in the first year of the 20-year period, 1997, and the highest percentage can be found in 2009 at 44.91%. This percentage of focus almost doubled from the lowest to the highest point. The largest annual percentage growth can be found during 1999 to 2000, which denotes an increase from 23.59% to 40.33%. The previously mentioned trend of an increase in the number of published articles in the beginning years and a leveling during the later years is duplicated in this percentage graph. This could indicate a growing interest in the topic of online learning in the early years when Internet technology was newer. Both graphs demonstrate a trend similar to the one found in Figure 3 reflecting Natividad's finding regarding online learning publication trends. Though interesting, it should be noted that this research includes articles related to online learning, and the other is from one factor within the broader interest of educational technology.

In order to look at publication trends on a more macro level, the data was grouped in multiyear periods. The final bibliometric graphic results in Figure 19 combines yearly online learning totals for five-year periods.

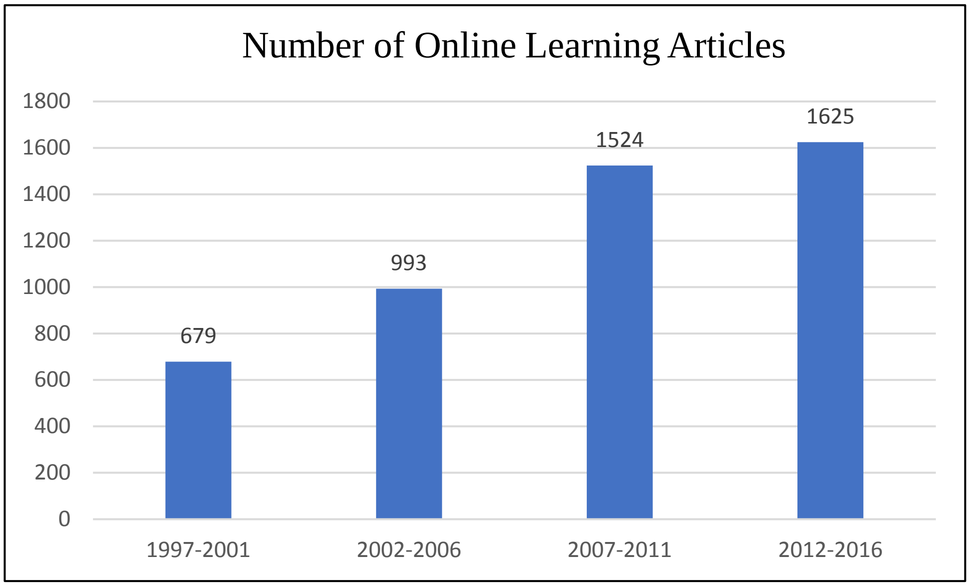

Figure 19. Number of online learning articles published in the 15 journals per 5-year period.

Each 5-year period shows an increase in the number of published articles. The largest difference is between the second and third periods of 531 articles. The next largest difference is between the first and second period with 334 articles. The smallest difference is evident between the final two time periods of 101 articles. These numbers correspond with the noted leveling of publication numbers in the latter years of the 20-year time span. The 5-year blocks were also used when examining the trends in terms used in online learning research.

Trends in Terms

The results for trends in terms address Research Question 2: What are the past, present, and emerging trends of terms regarding online learning within the selected literature during the past 20 years (1997-2016)?

The analysis was completed in four 5-year increments to obtain a clear picture of the trending terms throughout the entire 20-year period. Commonly used words have been filtered out of the abstract data. Trends for each five-year period are presented sequentially from the earliest to most recent. Stemmed n-grams of the abstract text served as the data source. In each period, a table providing information on the top 25 n-grams is given. Each table shows the number of times an n-gram is used and its average use per article. More exhaustive tables containing the 100 most frequent n-grams for each section are housed in Appendix D. Also presented in each block is a bar graph of the top 25 n-grams and a word cloud representing the top 100 words. Not included in the clouds are words on the stop word lists mentioned in the previous chapter. The final trending term analysis presented in this section provide results encompassing the entire 20-year time span.

Online Learning Research Term Trends 1997-2001

During 1997-2001, 679 articles were published in the 15 journals. The n-grams for the most popular meaningful terms used in online learning research articles during 1997-2001 are shown in the Table 7. The maximum number of terms per n-gram is three.

The top trending term, *learn* is used over 1,000 times with an average of 1.54 times per article abstract. One two-term n-gram, *distance_educ,* makes the top 25 list. The two corresponding stems, which comprise the n-gram, *distanc* and *educ*, are third and fifth on the list respectively. The top four stems—*learn, student, educ,* and *cours*—seem to relate more to the topic of learning in *online learning* rather than the more technology-oriented word *online*. The more technology-oriented terms of *web, technolog,* and *onlin* appear further down the list but are in the top 10 terms. A graphical depiction of this table is show in Figure 20.

Table 7

Top 25 N-Grams for Online Learning Article Abstracts (1997-2001)

Rank	N-Gram	Total	Per Article
1	learn	1046	1.54
2	student	945	1.39
3	educ	885	1.30
4	cours	641	0.94
5	distanc	523	0.77
6	web	504	0.74
7	base	493	0.73
8	technolog	453	0.67
9	develop	444	0.65
10	onlin	410	0.60
11	design	371	0.55
12	internet	348	0.51
13	commun	345	0.51
14	environ	309	0.46
15	comput	301	0.44
16	system	296	0.44
17	interact	287	0.42
18	instruct	281	0.41
19	distance_educ	280	0.41
20	teach	274	0.40
21	univers	256	0.38
22	inform	237	0.35
23	discuss	234	0.34
24	support	228	0.34
25	teacher	224	0.33

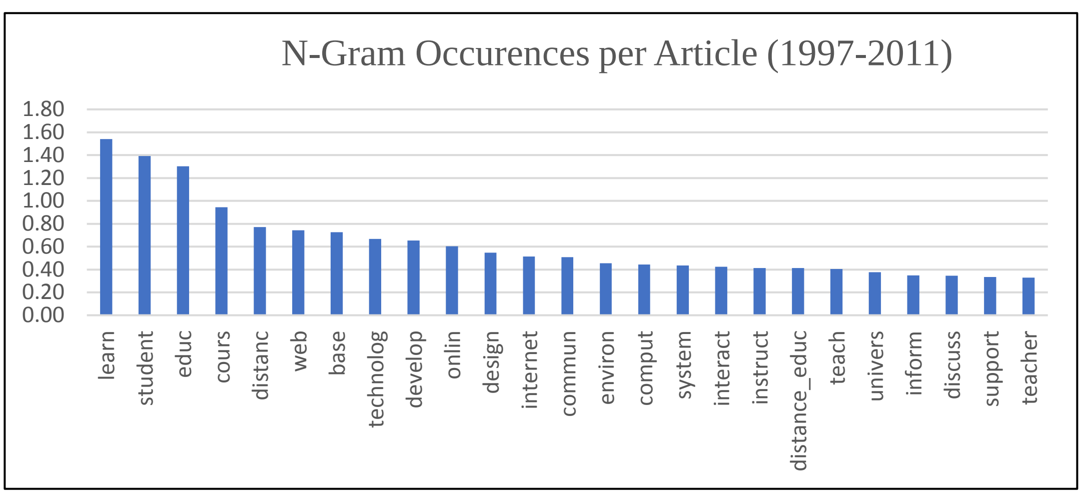

Figure 20. Graph depicting n-gram occurrences per article (1997-2001).

Interesting to note is that the more popular stems of *learn, student,* and *educ* average over one occurrence per article and the stems of *discuss, support,* and *teacher* appear in approximately one third of the articles. Per article, the most popular stems are used over three times as frequently as the lower three terms in the top 25 list. Another manner in which to view the data and recognize trends is with a picture of the actual words used in the article abstracts, a word cloud.

A word cloud containing the 100 most frequently used words in the abstracts of online learning articles is shown in Figure 21. The words correspond with the stemmed n-grams in the table and graph. The more times a word is used in a document, the larger the font.

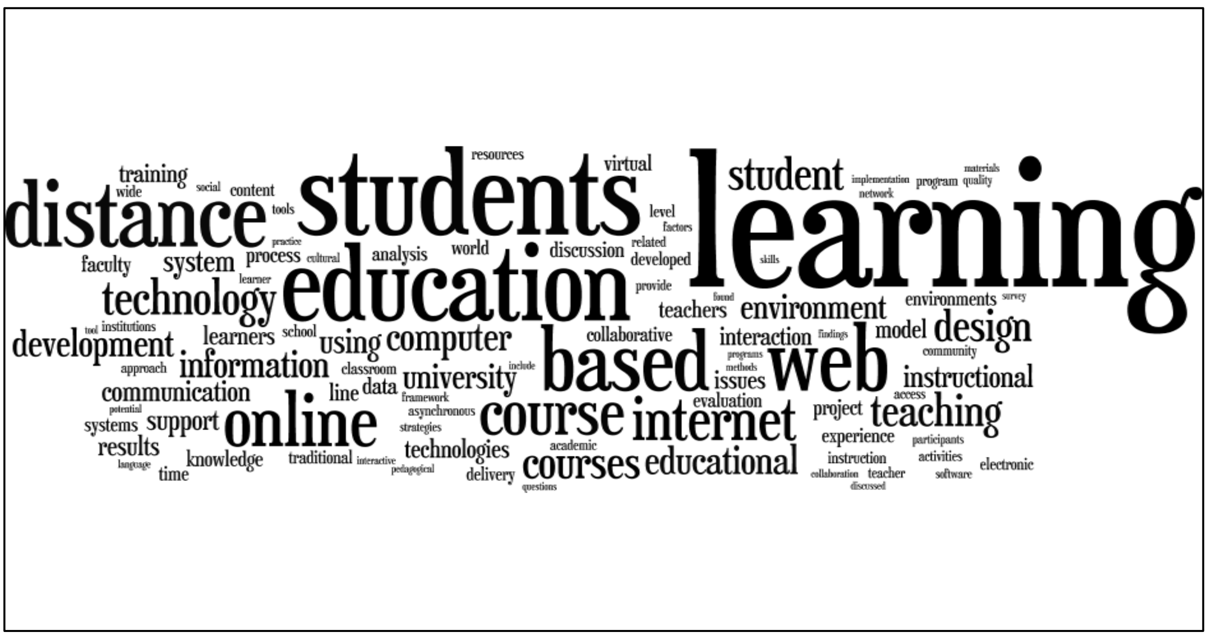

Figure 21. Word cloud for online learning article abstracts (1997-2001).

The largest word in the word cloud is *learning* denoting that is the word use most often in the article abstracts in this group of data. Unlike the stemmed n-gram, *learn*, the word *learning* here is for that word and no other conjugations. To clarify, the words *students* and *student* both appear in the word cloud, but in the n-gram list, the term *student* is used to represent the aggregate use of the stem.

Online Learning Research Term Trends 2002-2006

The next time period (2002-2006) contains 993 online learning articles from the 15 journals. Several of the same terms make an appearance in the top 25 list with a few variations. However, the placement of many of the n-grams in the list is different. The top 25 n-grams for 2002-2006 are shown in Table 8.

Table 8

Top 25 N-Grams for Online Learning Article Abstracts (2002-2006)

Rank	N_Gram	Total	Per Article
1	learn	2382	2.40
2	student	1839	1.85
3	onlin	1385	1.39
4	educ	983	0.99
5	cours	973	0.98
6	base	781	0.79
7	develop	696	0.70
8	web	647	0.65
9	design	597	0.60
10	technolog	542	0.55
11	environ	496	0.50
12	discuss	494	0.50
13	commun	484	0.49
14	learner	475	0.48
15	distanc	471	0.47
16	support	434	0.44
17	instruct	426	0.43
19	interact	420	0.42
18	system	405	0.41
20	teacher	391	0.39
21	collabor	389	0.39
22	effect	381	0.38
23	model	372	0.37
24	univers	364	0.37
25	teach	358	0.36

The top two n-grams of *learn* and *student* remain unchanged in ranking, but the frequency per article they are used has changed. New terms on the list include *learner, collabor, effect,* and *model.* N-grams that are no longer included in the top 25 are *inform, internet, distance_educ,* and *comput.* The term *onlin* is now the third most popular, up from tenth in the previous list. Other terms that made a five-step jump or greater (up or down) include: *distanc* (5 to 15), teach (20 to 25), *discuss* (23 to 12), *support* (24 to 16), *and teacher* (25 to 20). The graph depicting per article trend for the use of n-grams is shown in Figure 22.

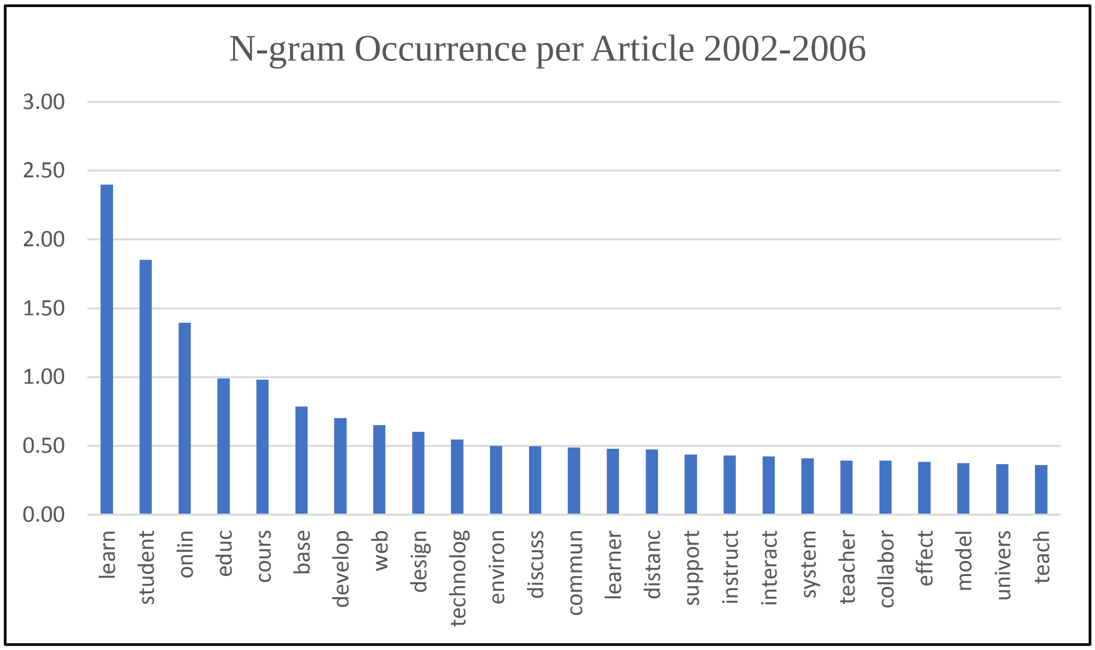

Figure 22. Graph depicting n-gram occurrences per article (2002-2006).

Changes in word use per article from the previous time period include the top two n-grams. The n-gram stem *learn* is used, on average, 2.4 times per article, an increase from 1.54 in the previous time period. The second most used stemmed n-gram, *student*, is used an average of 1.85 times per article compared to 1.39 in the prior time period. No other term with a difference average use of at least .5 times appeared in both graphs,. However, the term *interact* continues with the same average of .42 as the previous time period.

A word cloud giving a pictorial representation of words used in the article abstracts published 2002-2006 was created. These words contain the stemmed terms displayed in the table and graph. The word cloud is found in Figure 23.

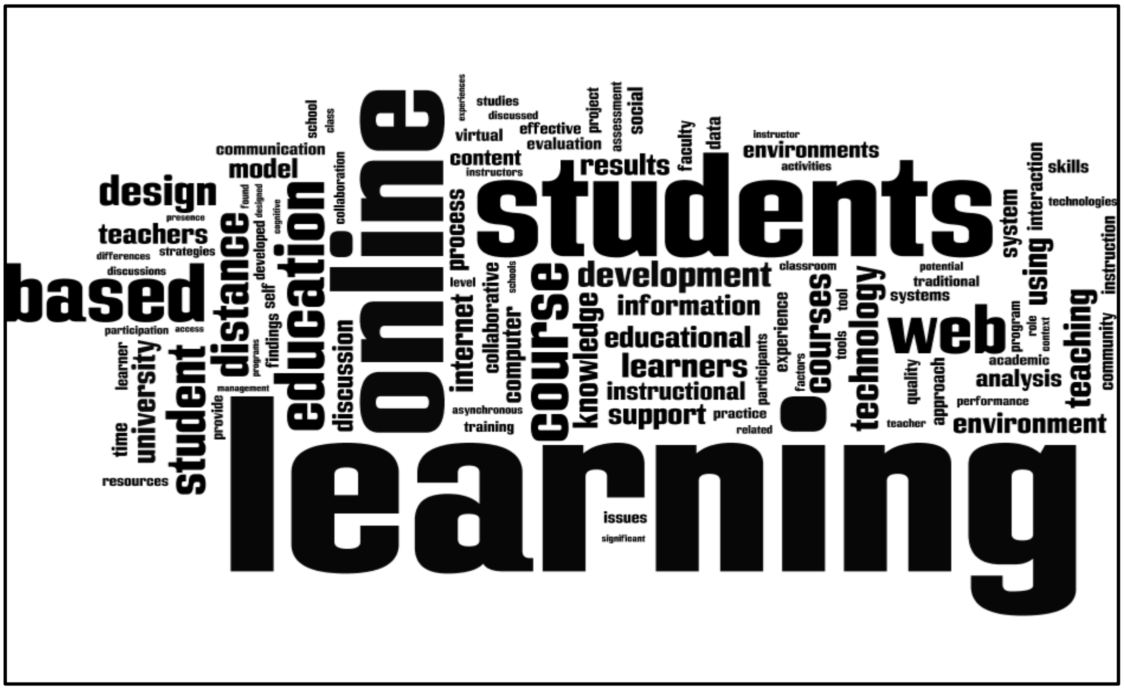

Figure 23. Word cloud for online learning article abstracts (2002-2006).

In the word cloud, the word *learning* continues to hold prominence as the largest word and therefore, occurred most often in the article abstracts. The word *students* also maintains a large presence in this cloud. Noted are the next three biggest words: *online, web,* and *based.* This could indicate an increased offering of online and Web-based courses in academia. The word *distance* appears smaller, which corresponds with its decline in the list and per article use of its stemmed n-gram.

Online Learning Research Term Trends 2007-2011

The initial time period in the second half of the data group contains 1,524 articles relating to online learning. Most terms remained close to the same average frequency use, but there was some movement on the top 25 list. The synopsis of n-gram use in these articles is shown in Table 9.

Table 9

Top 25 N-Grams for Online Learning Article Abstracts (2007-2011)

Rank	N-Gram	Total	Per Article
1	learn	4238	2.78
2	student	3332	2.19
3	onlin	2102	1.38
4	educ	1570	1.03
5	base	1103	0.72
6	cours	1087	0.71
7	develop	994	0.65
8	teacher	884	0.58
9	technolog	872	0.57
10	learner	863	0.57
11	design	851	0.56
12	result	800	0.52
13	system	800	0.52
14	web	749	0.49
15	environ	748	0.49
16	effect	729	0.48
17	discuss	722	0.47
18	support	670	0.44
19	interact	664	0.44
20	commun	638	0.42
21	teach	632	0.41
22	experi	607	0.40
23	model	605	0.40
24	particip	603	0.40
25	knowledg	602	0.40

The top seven terms—*learn, student, online, educ, base*, *cours,* and *develop*—remained in the top seven from the previous time period. The place of these terms remained the same, except for *base* (5th) and *cours* (6th), which switched places. New stemmed singular n-grams added to the top 25 list were *experi*, *particip*, and *knowledge.* Terms that exited the list include *collabor, effect,* and *univers.* Four terms moved five or more places within the list. Moving up on the list were *teacher* (20 to 8) and *system* (18 to 13). Moving down the list were *web* (8 to 14) and *discuss* (12 to 17). This movement of terms can be seen with the occurrences per article used graphic, shown in Figure 24.

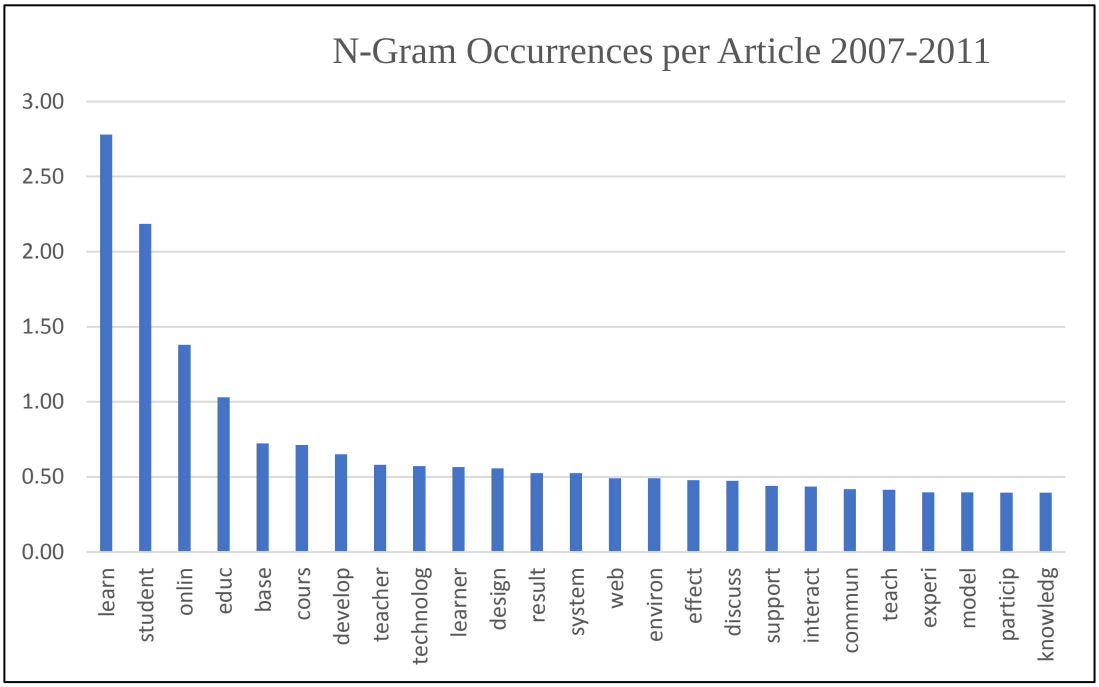

Figure 24. Graph depicting n-gram occurrences per article (2007-2011).

Within the top 25 terms, the top two had the greatest increase in per article occurrence. The terms *learn* (2.4 to 2.78) and *student* (1.85 to 2.19) demonstrated over a .3 increase. No other terms within the graphs showed an increase or decrease over .3 average occurrences. One term, *support*, remained constant at .44, compared to the prior time period. The graph depicts the nearly seven-fold increase between the average of the last term on the list, *knowledge* (.4) and the average use of the first term *learn* (2.78). Unchanged, thus far, within all three time periods are the first-ranked term, *learn*, and the second-ranked term, *student*. The next graphic, a word cloud in Figure 25, demonstrates the difference of these and more terms using the actual words within the abstracts.

Figure 25. Word cloud for online learning article abstracts (2007-2011).

Shown in Figure 25 are the top 100 words used in the online learning article abstracts for the years 2007-2011. The three most popular words in this time period—*learning*, *students*, and *online*—appear in the largest size font, which relates to the uses of their respective stems in the article abstracts. The word *learning* maintains the largest presence in the word cloud, which is consistent with the previous two word clouds. The word *education* and *base* are somewhat larger than the rest of the words in the cloud.

Online Learning Research Term Trends 2012-2016

The fourth and final time period contains 1,625 articles. A shift in the terms at the top of the list appears. The top 25 singular n-grams terms for this time period is in Table 10.

Table 10

Top 25 N-grams for Online Learning Article Abstracts (2012-2016)

Rank	N_Gram	Total	Per Article
1	student	4336	2.67
2	learn	4136	2.55
3	onlin	2930	1.80
4	educ	1746	1.07
5	cours	1624	1.00
6	develop	1050	0.65
7	design	1034	0.64
8	result	1005	0.62
9	base	1004	0.62
10	technolog	908	0.56
11	particip	898	0.55
12	effect	882	0.54
13	teacher	844	0.52
14	learner	826	0.51
15	discuss	766	0.47
16	environ	760	0.47
17	interact	691	0.43
18	social	689	0.42
19	teach	688	0.42
20	support	653	0.40
21	univers	630	0.39
22	activ	629	0.39
23	system	624	0.38
24	commun	616	0.38
25	inform	614	0.38

A change in the top two terms occurred for the first time during this final time period. The terms *student* and *learn* switched places. Taking the first position of most frequently used stemmed term is *student*, and *learn* is now in the second spot. Terms moving five places or more in the rankings include *particip* (24 to 11) and *system* (13 to 23). Four singular n-grams were deleted from the top 25 list, *experi*, *model*, *knowledge*, and *web*. These terms were in three of the four lowest slots for the previous time period, except for *web*, which was ranked at 14 during the previous round. Taking the places of the four deleted terms for 2012-2016, were four new terms—*social, univers, active,* and *inform.* The term *univers* appeared initially on the list for years 1997-2001 and the second list 2002-2006 but did not place in the top 25 for the third time period and appeared again for the current time period of 2012-2016. Another way to view the data is with the average use per article, which is shown numerically in Table 10 and graphically in Figure 26.

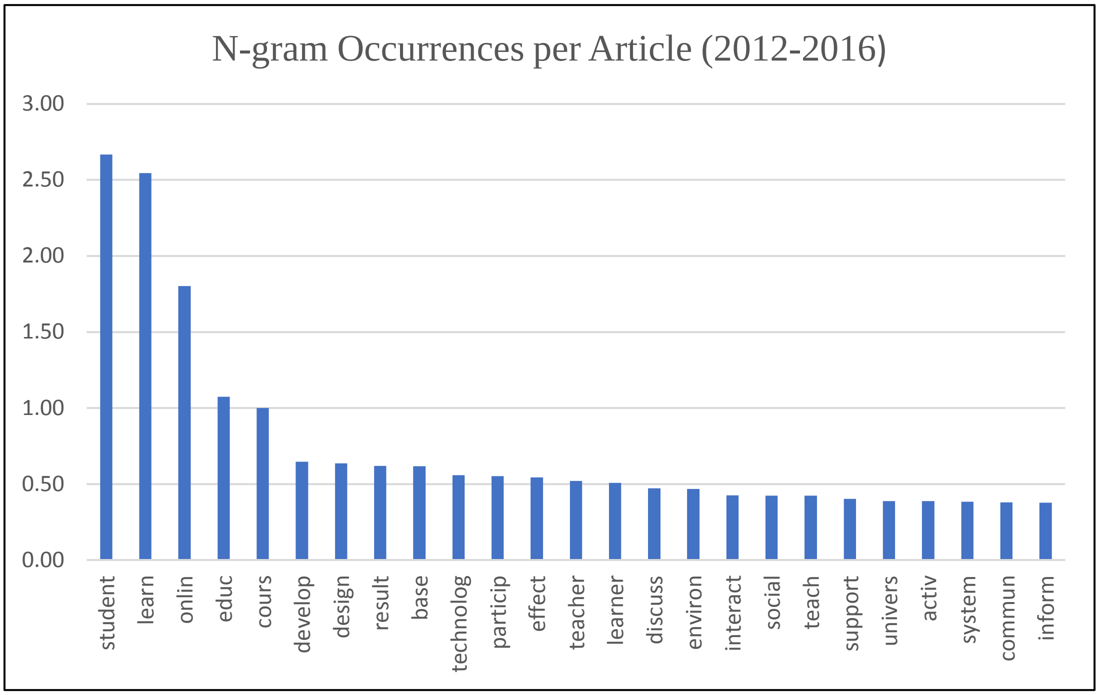

Figure 26. Graph depicting n-gram occurrences per article (2012-2016).

Increases greater than .5 occurred for the term *student* (2.19 to 2.67). In fact, *student* appears over seven times more than *inform*, the final term on the list. No other term experienced a difference of .5 or greater. However, the term *online* (1.38 to 1.8), experienced the second most average per-article gain. Terms that maintain generally the same per-article use are *develop* (.65) and *discuss* (.47). The term *technolog* (.57 to .56) had a .01 difference, but due to rounding could possibly be as close or closer than the other two terms. A word cloud composed of the 100 most frequently used words based on these stems is shown in Figure 27.

Figure 27. Word cloud for online learning article abstracts (2012-2016).

Figure 28 displays the 100 most frequently occurring words found in online learning article abstracts for the years 2012-2016. The larger fonts of the words *student* and *students* correspond to the term *student* as the most frequently used stem. The three largest words —*students, learning,* and *online*—remain constant from the two previous word clouds for the time periods 2002-2006 and 2007-2011. In the word cloud for 2012-2016 shown in Figure 28, the words *based* and *education* appear to be the next largest terms. In this word cloud, other words of similar size to these two are *course, results,* and *student.*

Online Learning Research Term Trends 1997-2016

Combining all time periods provides an overview of term usage in online learning abstracts throughout the 20-year period. The 25 most frequently used stemmed n-grams are shown in Table 11.

Table 11

Top 25 N-Grams for Online Learning Article Abstracts (1997-2016)

Rank	N-Gram	Total	Per Article
1	learn	11802	2.45
2	student	10452	2.17
3	onlin	6827	1.42
4	educ	5184	1.08
5	cours	4325	0.90
6	base	3381	0.70
7	develop	3184	0.66
8	design	2853	0.59
9	technolog	2775	0.58
10	web	2420	0.50
11	learner	2382	0.49
12	result	2352	0.49
13	teacher	2343	0.49
14	environ	2313	0.48
15	discuss	2216	0.46
16	effect	2161	0.45
17	system	2125	0.44
18	distanc	2106	0.44
19	commun	2083	0.43
20	interact	2062	0.43
21	particip	1991	0.41
22	support	1985	0.41
23	teach	1952	0.40
24	model	1778	0.37
25	univers	1768	0.37

This table represents the top stemmed n-grams for words found in the 4,821 online learning article abstracts published in the years 1997-2016. The stem *learn* was the most frequently used at 11,802 times. Terms that have consistently been ranked in the top 25 frequently used terms include *learn*, *student*, *onlin*, *educ*, *cours*, *base*, *develop*, *design*, *technolog*, *teacher*, *environ*, *discuss*, *system*, *commun*, *interact*, *support*, and *teach.* These 17 terms could represent stable vocabulary and topics appearing in online learning articles across the two decades. Another measurement is the average use of the terms per article. The following image, Figure 28, depicts this term value.

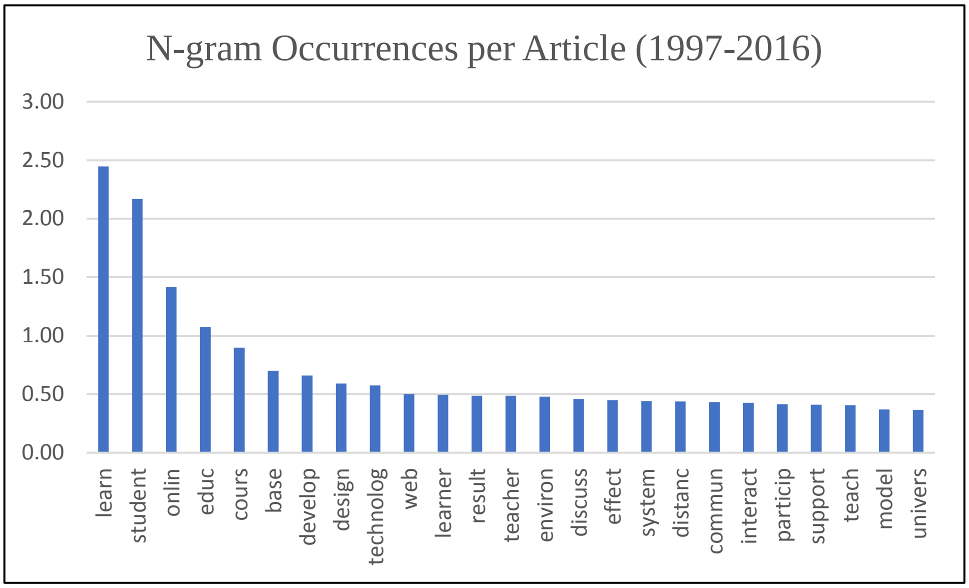

Figure 28. Graph depicting n-gram occurrences per article (1997-2016).

In the graph displayed in Figure 28, the top terms boast an average of over two occurrences per article abstract. The top term, *learn* is used over six times more per article than the term with the least frequent average on the list, *univers*. All five graphs displaying the per-article use, show a similar steeped curve with the first few terms being discernably higher, then the graph appears to become more level at the fifth term. Words comprised using these terms are seen in the word cloud in Figure 29.

Figure 29. Word cloud for online learning article abstracts (1997-2016).

In the word cloud, the font size directly relates to the frequency of word use. The word *learning* is the largest word because it is the most frequently used word in all of the 4,821 online learning article abstracts. The word cloud represents the 100 most frequently used words in online article abstracts throughout two decades (1997-2016).

Cluster Results

Examining the trends in terms provides a view of the development and movement of research topics from the standpoint of the language used in the text of the article abstracts. Another view is to investigate categories of research within the two decades. Results in this section present a flat categorical taxonomy of the articles' abstract text and address the third and research question: What are the thematic clusters of the articles regarding online learning in the 15 journals for the past 20 years (1997-2016)?

Cluster Overview

The clusters were created based on the output from the RapidMiner (2017) k-means cluster program. An Excel file containing 4,821 rows of data, each representing a single online learning article from the 15 journals. To obtain the clusters, a k-means cluster analysis was performed. Eighteen unique groups were identified and labeled. Each cluster label is based on the (a) highest weighted terms, (b) article titles, and (c) contents of the article abstracts. Table 12 gives basic information on the identified clusters.

Table 12

Overview of Cluster Categories

Cluster ID	Cluster Label	# of Articles	% of Articles	Mean Year	Run #
C1	Web Issues	327	6.78%	2004.6	11
C2	Distance Education	314	6.51%	2005.8	17
C3	Internet	197	4.09%	2006.3	10
C4	Technology and Education	976	20.24%	2007.5	4
C5	Faculty Concerns	225	4.67%	2008.1	14
C6	Learner Considerations	268	5.56%	2008.2	3
C7	Collaboration	241	5.00%	2008.5	12
C8	Discussion Forums and Communication	273	5.66%	2008.7	5
C9	Virtual Environment and Simulations	211	4.38%	2008.8	0
C10	Teaching	246	5.10%	2009.0	6
C11	Journal Articles	41	0.85%	2009.4	16
C12	Multimedia	50	1.04%	2009.4	2
C13	Student Assessment	900	18.67%	2009.7	8
C14	Learner Motivation and Efficacy	156	3.24%	2010.4	1
C15	Social Presence	152	3.15%	2010.4	7
C16	Openness and Mobility	105	2.18%	2010.4	9
C17	Game-Based Learning	71	1.47%	2011.6	13
C18	MOOCs	68	1.41%	2014.8	15
	Totals	4,821	100.00%		

In order to quickly distinguish each cluster, an exclusive identification code was assigned. The numerical order of the clusters sequentially follows the average publication year of the articles in that cluster.

- The first column of the tables shows the cluster ID.
- The second column is the cluster label, also referred as the cluster name.
- The third column indicates how many articles are assigned to the cluster. The numbers in this column total 4,821, indicating that all articles included in this study are assigned to a cluster.
- The fourth column provides the weight of that cluster given in percentage form, based on the total number of articles used in this study (4,821). Note that the percentages total 100%, providing another verification check that all articles are assigned to a cluster.
- The fifth column is the average year of publication for the articles in the cluster.
- The final column represents the number given that cluster at the end of the clustering program run.

The table offers some interesting results. The largest cluster, C4: Technology and education, is not totally unexpected because 10 of the 15 journals used as data sources were used in a prior study focused on educational technology (Natividad, 2016) . It also represents an intersection of online learning with learning technologies and emphasizes their connection. The second largest cluster, C13: Student assessment, could be surprising as it is a topic that is not specifically related to online learning, but to education as a whole. These two clusters, comprising over 40% of the data, could be considered relevant academic topics outside of online education.

One cluster that is surprising given its relatively small size compared with the other clusters is C12: Multimedia. The internet and digital technology lends a backdrop on which various forms of multimedia educational content can be disseminated. Perhaps some of the articles which mention multimedia aspects of online learning were absorbed into other clusters. Another small cluster that could be unexpected is C18: MOOCs. Given the recent popularity of this topic, a larger cluster may have been anticipated. However, since this is a relatively new interest in the distance education world, its research history does not date back 20 years. To give a more extensive view of the resulting clusters, a more detailed examination into each of the clusters is warranted.

Cluster Trends

The results from the time trend analysis address the fourth research question: What are the time trends of the recognized thematic clusters found in the selected articles during the past 20 years (1997-2016)? The analysis yielded six types of trends within the clusters: (a) decreasing, (b) steady, (c) peaked, (d) fluctuating, (e) increasing, and (f) emerging. This section shows tables that list the highest weighted centroid terms for each cluster (see Tables 13-18). Each term in the table has a weighted value based on the centroid and tf-idf value calculations of at least 0.02, and terms are listed in descending order. The higher the value, the closer the term's location to the center of the cluster. Following each table, a discussion for those clusters is provided. Within the discussion for each cluster, the following is conferred: (a) five highest weighted terms, (b) three example article titles, and (c) dialogue about the cluster time trends accompanied by a graphical depiction. Further details about each of the clusters can be found in Appendix E.

Decreasing Clusters

The first three clusters demonstrate a general trend decline over the twenty-year period. The highest weighted stems for these clusters is found in Table 13. While the topics may remain pertinent in online learning research, a decrease in their use within the research is noted.

Table 13

Centroid Table with Weighted Attribute Values for Declining Clusters (C1 – C3)

C1 Web Issues		C2 Distance Education		C3 Internet	
web	0.158	distanc	0.18	internet	0.225
site	0.059	educ	0.065	search	0.08
base	0.049	cours	0.035	parent	0.048
tool	0.042	technolog	0.03	school	0.038
system	0.042	univers	0.028	inform	0.038
design	0.035	student	0.025	comput	0.03
instruct	0.035	institut	0.025	attitud	0.028
inform	0.03	transact	0.021	student	0.028
cours	0.028	learner	0.021	children	0.025
technolog	0.027	deliveri	0.02	usag	0.025
develop	0.026	learn	0.02	anxieti	0.022
user	0.025	issu	0.02	access	0.02
page	0.024				
evalu	0.023				
model	0.023				
wide	0.023				
student	0.022				
resourc	0.022				
educ	0.021				
learn	0.02				

C1: Web issues. This cluster contains 327 articles accounting for 6.78% of all online learning articles. The top five stemmed terms in this cluster are *web, site, base, tool,* and *system.* The theme for this cluster centers around topics related to Web sites and the World Wide Web. Example titles included in this cluster are (a) “Learning with the Web: Experience of Using the World Wide Web in a Learning Environment” (Sloane, 1997), (b) “A New Tool for Managing

Students' Questions in Web-Based Distance Education Courses" (Scapin & Marega, 2000), and (c) "A Web-Based Formative Assessment Tool for Masters Students: A Pilot Study" (Costa, Mullan, Kothe, & Butow, 2010). The average publication year for articles is 2005 and the median publication year is 2003 with most of the articles, 37, published in 2000. When looking at percentages of all articles published per year, as seen in Figure 30, a general overall decline is depicted.

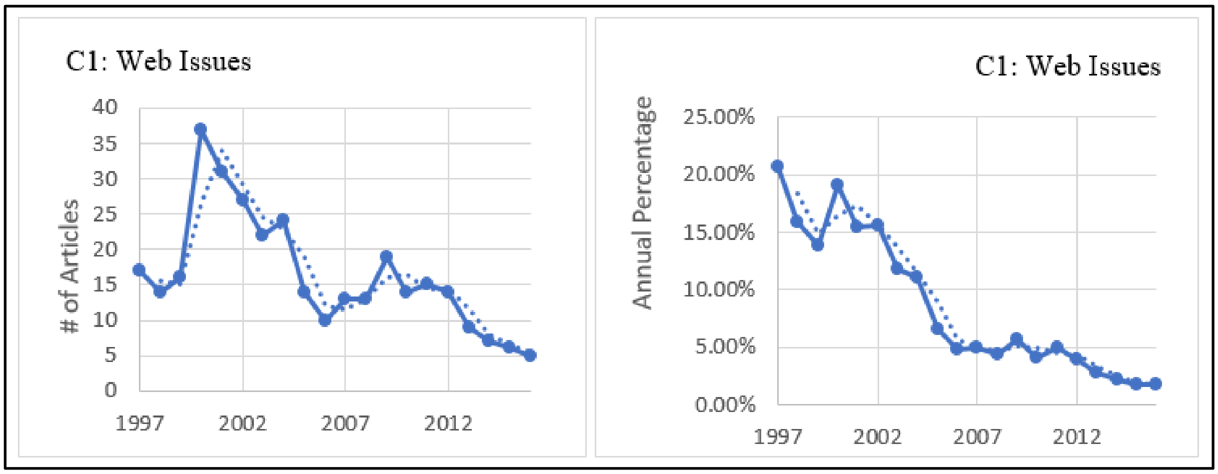

Figure 30. Graphs depicting article distribution in C1: Web Issues.

C2: Distance education. Housed in this cluster are 314 online learning articles, which comprise 6.51% of the articles in this study. Article topics focus on distance education and its relationship to online learning. The most highly weighted terms for this cluster are *distanc*, *educ*, *cours*, *technolog*, and *univers*. As mentioned previously, distance education is a broad topic that can encompass online learning. Example articles in the cluster include (a) "A Case Study of Technology use in Distance Learning" (Zhang, 1998) (b) "Pedagogical Implications of Working with Doctoral Students at a Distance" (Wikeley & Muschamp, 2007), and (c) "Student Retention in Distance Education: Are We Failing Our Students?" (Simpson, 2013). The median publication years for articles in this cluster are 2005 and 2006, with the most articles published

(25) in the years 2000 and 2001. Though the number of articles published yearly remains steady, the yearly percentage generally declines through the 20-year period. Figure 31 displays the yearly trend graphs.

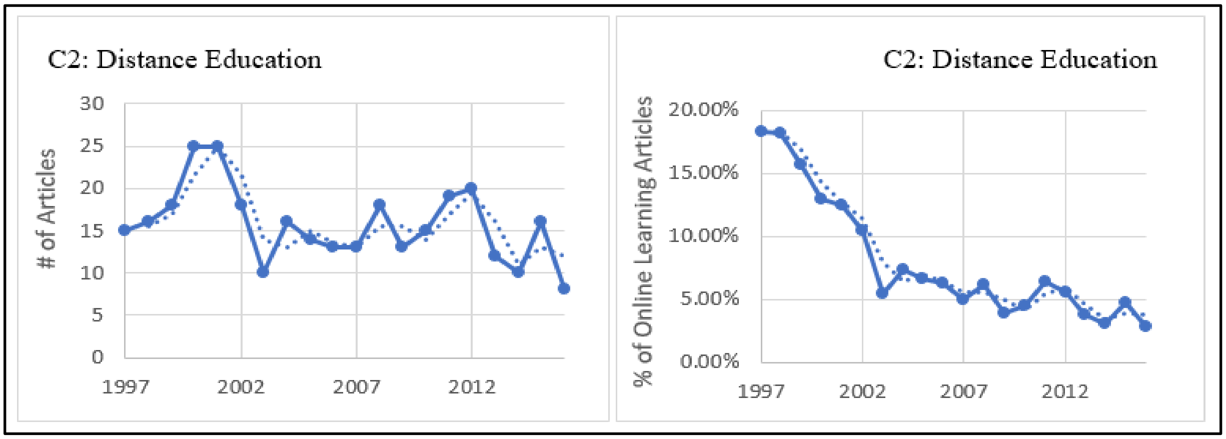

Figure 31. Graphs depicting article distribution in C2: Distance Education.

C3: Internet. This cluster contains 197 articles focused on topics pertaining to the Internet. The number of articles in this study makes up 4.09% of the researched documents. The top five weighted terms are *internet*, *search*, *parent*, *school*, and *inform*. Articles include (a) "The Effects of Online Multimedia Project Development, Learning Style, and Prior Computer Experiences on Teachers' Attitudes toward the Internet and Hypermedia" (Takacs, Reed, Wells, & Dombrowski, 1999), (b) "Online Information Searching Strategy Inventory (OISSI): A Quick Version and a Complete Version" (M.-J.Tsai, 2009), and (c) "Three Questions about the Internet of Things and Children" (Manches, Duncan, Plowman, & Sabeti, 2015). The year 2007 serves as the median for yearly publications, with the average being the year before, 2006. More articles, 18, were published in 2001 than any other year. The yearly number of published articles fluctuates, while the percentage per year demonstrates a steady decline after 1999. The graphs in Figure 32 depict these trends.

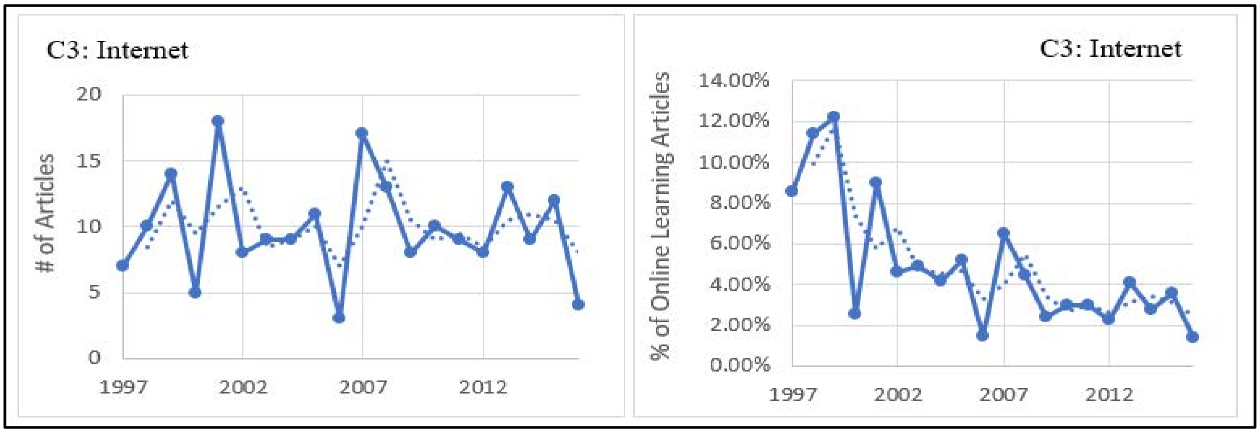

Figure 32. Graphs depicting article distribution in C3: Internet.

Steady Clusters

The largest cluster, C4: Technology and education along with C5: Faculty concerns remained fairly stable in the research literature over the two decades. Annual fluctuations were apparent, but the overall trend continued to be fairly steady. The centroid information for these two clusters is found in Table 14. Also presented is a more detailed discussion of the individual clusters.

C4: Technology and Education. The Technology and Education group makes up the largest of the clusters with 976 articles or 20.24%. Top-trending stemmed terms for this cluster include *technolog*, *educ*, *learn*, *institut*, and *develop.* The theme centers around technology and its role in education. Article titles contained in this cluster include (a) “Can New Technology Remove Barriers to Work-Based Learning?” (Reeve, Gallacher, & Mayes, 1998), (b) “A Personalisable Electronic Book for Video-Based Sign Language Education” (Ohene-Djan, Zimmer, Gorle, & Naqvi, 2003), and (c) “Networked Participatory Scholarship: Emergent Techno-Cultural Pressures Toward Open and Digital Scholarship in Online Networks” (Veletsianos & Kimmons, 2012). A graphical depiction of yearly publication data for this cluster is displayed in Figure 33. The mean and median year for publications of articles in this cluster is 2008. Seventy-one articles were published in 2010, the most prolific year for this cluster.

Table 14

Centroid Table with Weighted Attribute Values for Stable Clusters (C4 and C5)

(C4) Technology and Education		(C5) Faculty Concerns	
technolog	0.033	faculti	0.166
educ	0.029	program	0.135
learn	0.027	cours	0.053
institut	0.026	onlin	0.048
develop	0.023	teach	0.04
onlin	0.022	univers	0.037
project	0.022	colleg	0.035
univers	0.022	degre	0.034
system	0.021	student	0.031
commun	0.021	develop	0.03
design	0.021	institut	0.027
cours	0.020	educ	0.025
		doctor	0.025
		offer	0.024
		administr	0.024
		distanc	0.023
		satisfact	0.022
		graduat	0.022
		profession	0.021

C5: Faculty Concerns. This cluster includes 225 articles, which comprise 4.67% of the selected articles. The most frequently used words of the articles in this cluster contain the stems of *faculty, program, cours, onlin,* and *teach.* The topics of the article abstracts point to faculty involvement with online education. Examples of articles located in the cluster include: (a) "Factors Influencing Faculty Satisfaction with Asynchronous Teaching and Learning in the

SUNY Learning Network" (Fredericksen, Pickett, Shea, Pelz, & Swan, 2000), (b) "Motivators and Inhibitors for University Faculty in Distance and E-Learning" (Cook, Ley, Crawford, & Warner, 2008), and (c) "Developing a Quality Improvement Process to Optimize Faculty Success" (Merillat & Scheibmeir, 2016).

The median year for article publication is 2009. Twenty-four articles were published during the most prolific year, 2010. Graphs depicting yearly article publication are shown in Figure 34. The number of articles published fluctuates yearly, but have generally increased over the 20-year period. The percentage of article published each year has fluctuated as well, but the trend line depicts general stability over the two decades.

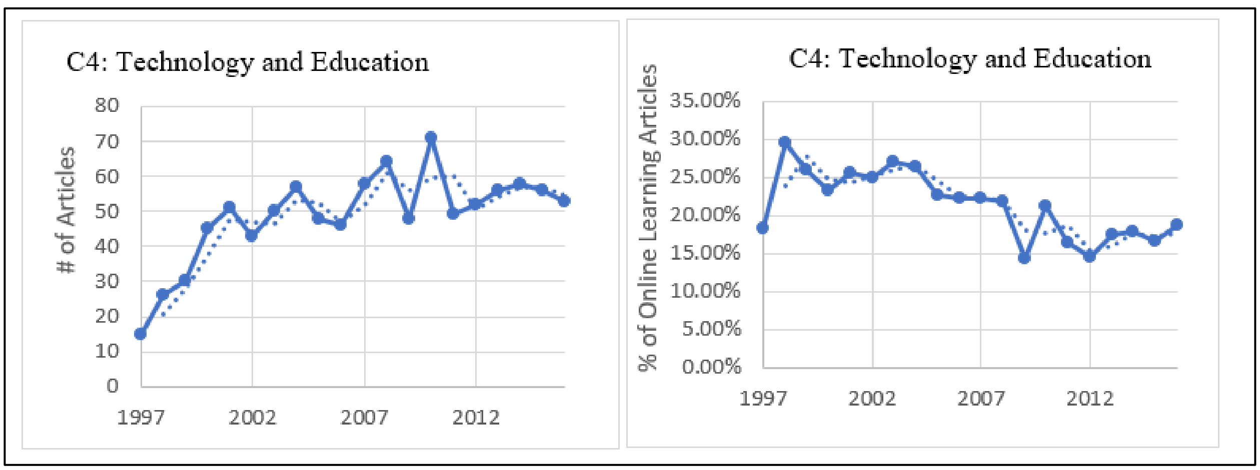

Figure 34. Graphs depicting article distribution in C4: Technology in Education.

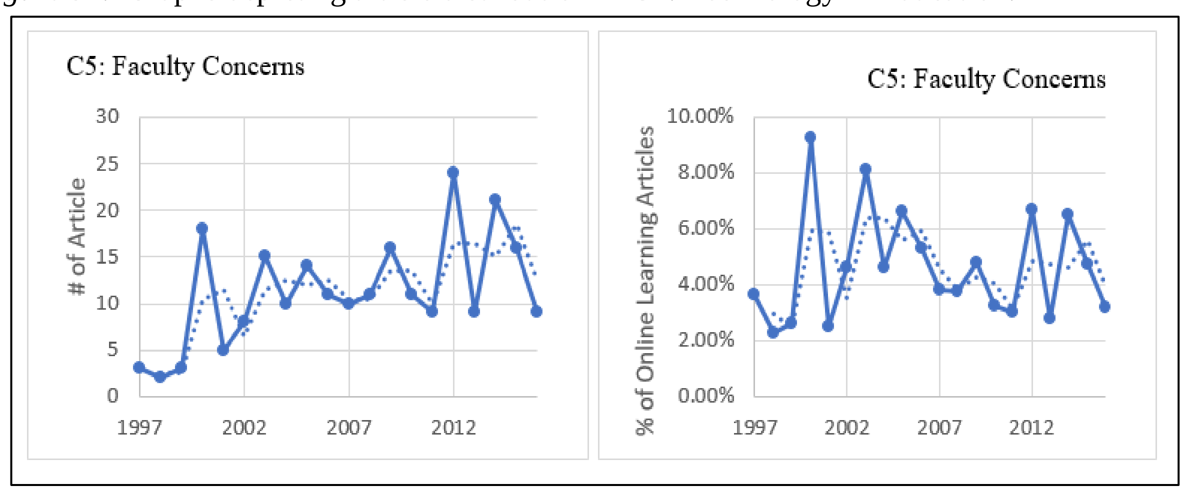

Figure 33. Graphs depicting article distribution in C5: Faculty Concerns.

Peaked Clusters

The clusters demonstrated perceptible consistent fluctuations throughout the two decades. However, a general overall increase followed by a decrease in article publication was noticed. The weighted terms are displayed in Table 15 followed by a discussion of individual cluster.

Table 15

Centroid Table with Weighted Attribute Values for Peaked Clusters (C6 – C8)

(C6) Learner Considerations		(C7) Collaboration		(C8) Discussion Forums and Communication	
learner	0.136	collabor	0.182	discuss	0.144
style	0.076	knowledg	0.077	forum	0.08
learn	0.048	share	0.041	asynchron	0.056
blog	0.041	commun	0.037	messag	0.055
system	0.037	activ	0.037	onlin	0.052
interact	0.035	interact	0.037	post	0.050
instruct	0.031	construct	0.036	particip	0.048
onlin	0.030	environ	0.035	student	0.043
satisfact	0.029	learn	0.035	instructor	0.040
model	0.029	onlin	0.032	interact	0.037
adult	0.029	support	0.031	facilit	0.035
design	0.026	student	0.029	analysi	0.032
cours	0.024	task	0.027	cours	0.032
reflect	0.023	build	0.026	discours	0.026
adapt	0.023	process	0.025	commun	0.026
support	0.023	cscl	0.024	thread	0.025
environ	0.023	comput	0.024	level	0.024
base	0.022	team	0.024	activ	0.024
perform	0.022	design	0.023	synchron	0.022
cognit	0.021	social	0.023	cognit	0.021
particip	0.021	individu	0.023	class	0.021
factor	0.021	tool	0.023	board	0.021
propos	0.020	particip	0.022	content	0.021
person	0.020	learner	0.021	pattern	0.020
effect	0.020	approach	0.021		
		teacher	0.021		
		base	0.021		
		web	0.02		

C6: Learner considerations. The sixth cluster contains 268 articles and centers on learners in an online education environment. This cluster houses 5.56% of the online learning journal articles. Comprising the top five terms in Learner Considerations are *learner*, *style*, *learn*, *blog*, and *system*. A few of the article titles in this cluster include (a) "Predictors of Learner Satisfaction in an Academic Computer Conference" (Gunawardena & Duporne, 2000), (b) "Effects of High Level Prompts and Peer Assessment on Online Learners' Reflection Levels" (Chen, Wei, Wu, & Uden, 2009), and (c) "Constructing Proxy Variables to Measure Adult Learners' Time Management Strategies in LMS" (Jo, Kim, & Yoon, 2015). The average publication year is 2008 and the median year for published articles is 2009. The number of articles published peaked in 2009 at 28. Graphs depicting yearly article publication data are presented in Figure 35. The number of articles showed a general increase, peaking in 2009, followed by a general decline. Though the past couple of years in this cluster show an increase, a similar trend follows in the yearly percentage of articles, although a dip appears between 2002 and 2007. The year of the largest percentage of published articles is 2002 at 9.25%.

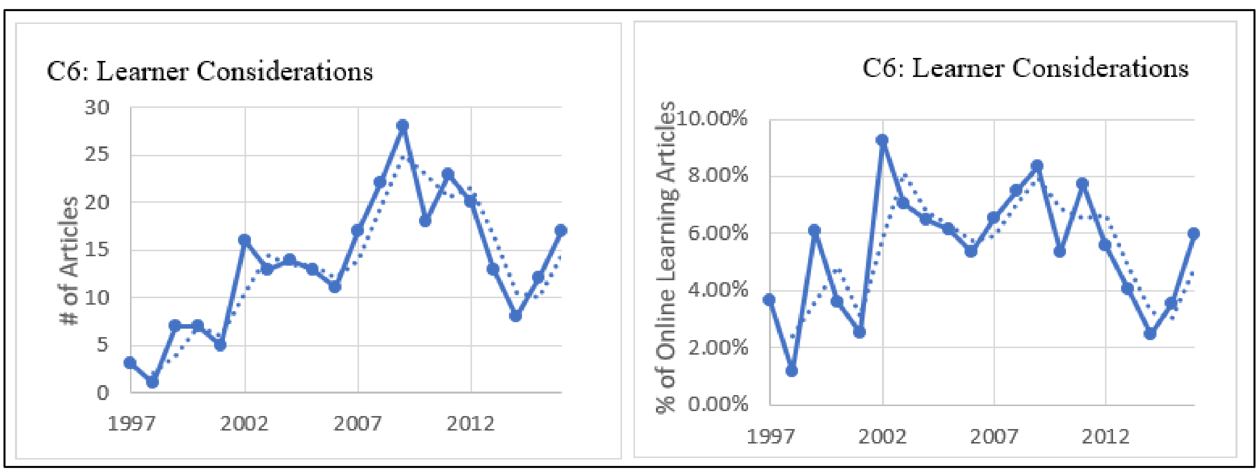

Figure 35. Graphs depicting article distribution in C6: Learner Considerations.

C7: Collaboration. The collaboration cluster houses 241 articles, or 5.00% of the total number of articles used in this research, with a focus on collaboration in online learning. The top five stemmed terms (in order) are *collabor, knowledg, share, commun,* and *activ.* Examples of titles in this cluster include (a) "Collaborative Learning via the Internet" (Ragoonaden & Bordeleau, 2000), (b) "Supporting Students to Develop Collaborative Learning Skills in Technology-Based Environments" (Nevgi, Virtanen, & Niemi, 2006), and (c) "Collaborative Argumentation and Cognitive Elaboration in a Computer-Supported Collaborative Learning Environment" (Stegmann, Wecker, Weinberger, & Fischer, 2012). Of the 241 articles, 22 were published in 2010, the most prolific year. The median publication year for the 20-year period is 2009, with the average being 2008. The article publication trend showed a general increase until 2010, after which fluctuations appear (see Figure 36). The percentage of online learning articles in this cluster show a similar trend; however, the peak percentage is in 2006.

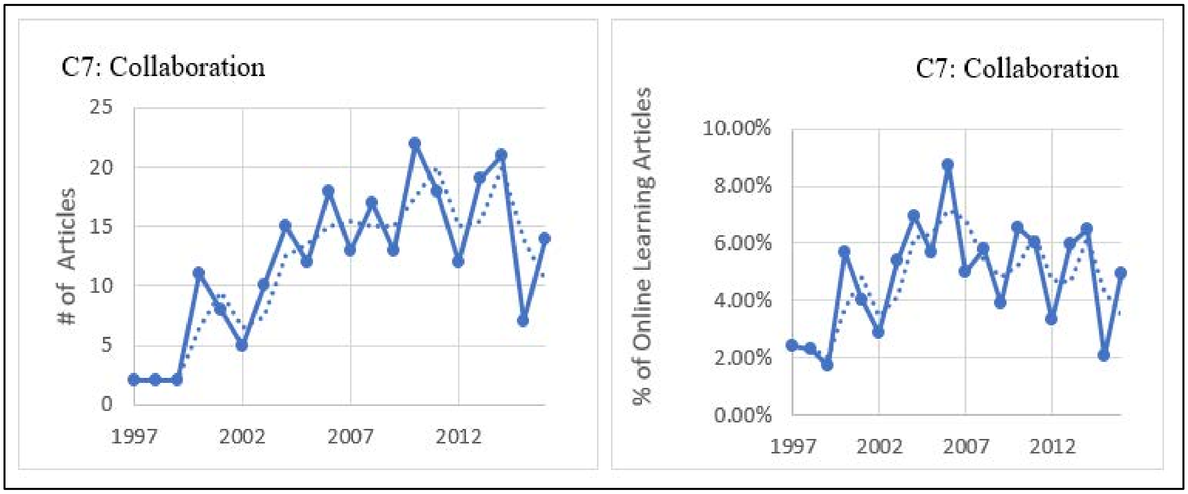

Figure 36. Graphs depicting article distribution in C7: Collaboration.

C8: Discussion forums and communication. This cluster maintains a focus on discussion forums and other means of communication and contains 241 articles, making up 5.66% of the body of article abstract data. The most used stems are *discuss*, *knowledg*, *share*, *commun*, and *interact*. Example articles in this cluster include (a) "Keeping Online Synchronous Discussions on Topic" (Beaudin, 1999), (b) "The Influence of the Discussion Leader Procedure on the Quality of Arguments in Online Discussions" (Spatariu, Hartley, Schraw, Bendixen, & Quinn, 2007), and (c) "Subject Line Preferences and other Factors Contributing to Coherence and Interaction in Student Discussion Forums" (Skogs, 2013). The mean and median publication year in this cluster is 2009. The year 2006 is the year with the most articles published, at 25. The number of articles and the percentage of articles published follow similar trends shown in Figure 37. A general increase occurs until 2006 and then the publications fluctuate. However, the last three years show a consistent decline.

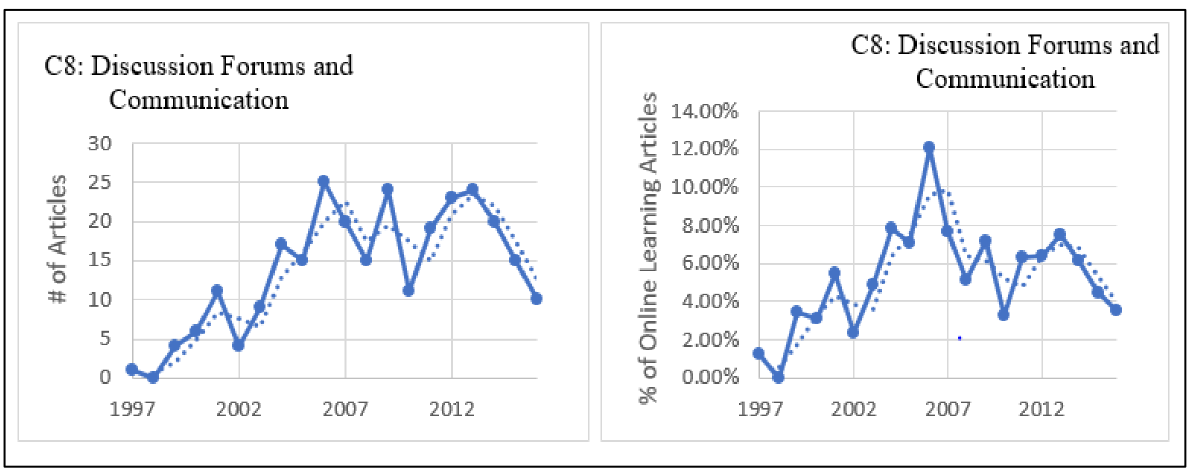

Figure 37. Graphs depicting article distribution in C8: Discussion Forums and Communication.

Fluctuating Clusters

The clusters belonging to this group largely showed yearly fluctuations throughout the 20 year time period. Due to the variability, a general trend was not noticed. The weighted terms for this group is found in Table 16. A discussion of the individual clusters is also given.

Table 16

Centroid Table with Weighted Attribute Values for Fluctuating Clusters (C9 – C12)

C9 Virtual Environment and Simulations		C10 Teaching and Instructional Delivery		C11 Published Research Information		C12 Multimedia	
virtual	0.198	teacher	0.204	journal	0.243	video	0.366
World	0.077	teach	0.05	articl	0.102	feedback	0.035
environ	0.049	profession	0.046	publish	0.087	digit	0.033
school	0.041	school	0.045	write	0.066	lectur	0.032
laboratori	0.037	preservic	0.04	paper	0.063	instruct	0.031
Reality	0.031	ict	0.035	review	0.055	student	0.029
Simul	0.031	practic	0.035	distanc	0.045	cours	0.026
Experi	0.027	onlin	0.034	citat	0.045	text	0.026
Learn	0.024	technolog	0.03	public	0.044	evalu	0.026
Life	0.024	develop	0.029	topic	0.04	stream	0.026
interact	0.024	classroom	0.027	trend	0.04	instructor	0.025
Design	0.024	interact	0.027	reflect	0.038	reflect	0.024
student	0.023	activ	0.026	educ	0.034	annot	0.023
User	0.023	educ	0.026	field	0.032	base	0.023
Muve	0.023	student	0.026	studi	0.032	multimedia	0.022
Educ	0.021	servic	0.026	methodolog	0.029	perform	0.022
		knowledg	0.025	aod	0.029	design	0.022
		commun	0.025	analysi	0.026	note	0.021
		pre	0.024	peer	0.023	self	0.021
		support	0.024	cite	0.023	effect	0.021
		train	0.024	forum	0.023	audio	0.020
		instruct	0.024	method	0.022	youtub	0.020
		pedagog	0.023	ajd	0.022	commun	0.020
		learn	0.022	period	0.021		
		project	0.022	ssci	0.021		
		particip	0.021	collabor	0.021		
		program	0.021	promin	0.021		
		role	0.021	issu	0.02		
		reflect	0.02				
		web	0.02				

C9: Virtual environments and simulations. This cluster is comprised of 211 articles focused on various virtual environments and digital simulations within online learning. Accounting for 4.38% of the articles, the top five weighted terms include *discuss, forum, asynchron, messag,* and *onlin*. Titles that make up this cluster consist of (a) "Internet Application of LabVIEW in Computer Based Learning" (Egarievwe, et al., 2000), (b) "Three-Dimensional Virtual Worlds and Distance Learning: Two Case Studies of Active Worlds as a Medium for Distance Education" (Dickey, 2005) and (c) "Technical Problems Experienced in the Transformation of Virtual Worlds into an Education Environment and Coping Strategies" (Coban, Karakus, Karaman, Gunay, & Goktas, 2015). The average publication year is 2009. The median and most prolific year is 2010, with 24 articles published, and is indicated graphically in Figure 38. However, the highest percentage of online learning articles in this cluster appeared in 1998. The trend of published articles fluctuates, peaking in 2010 and then declines. However, the final year of publication, 2016, does indicate slight growth that could signal the start of an increased research interest.

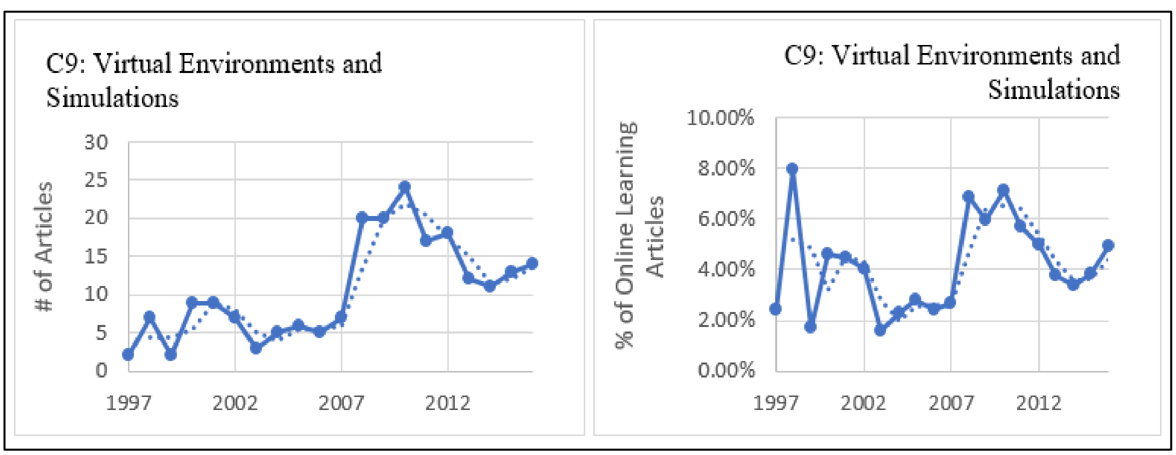

Figure 38. Graphs depicting article distribution in C9: Virtual Environments and Simulations.

C10: Teaching. The tenth cluster contains 246 articles related to teaching and online learning. Top stemmed terms for this group include *teacher*, *teach*, *profession*, *school*, and *preservice*. Example articles in this cluster include (a) "Determining the Impact of Training on Teacher Use for a Web-to-Database System" (S. Garrison, Fenton, & Vaissiere, 2001), (b) "Learning to Teach Online: What Works for Pre-Service Teachers" (Duncan & Barnett, 2009), and (c) "Does a University Teacher Need to Change E-Learning Beliefs and Practices When Using a Social Networking Site? A Longitudinal Case Study" (Scott, 2013). The mean and median publication year is 2009. That year also serves as the most prolific with 29 articles published. The article publication trend, as displayed in Figure 39, shows a general increase until the peak year of 2009, and then an overall decline follows. The trend is similar for the percentage of published articles.

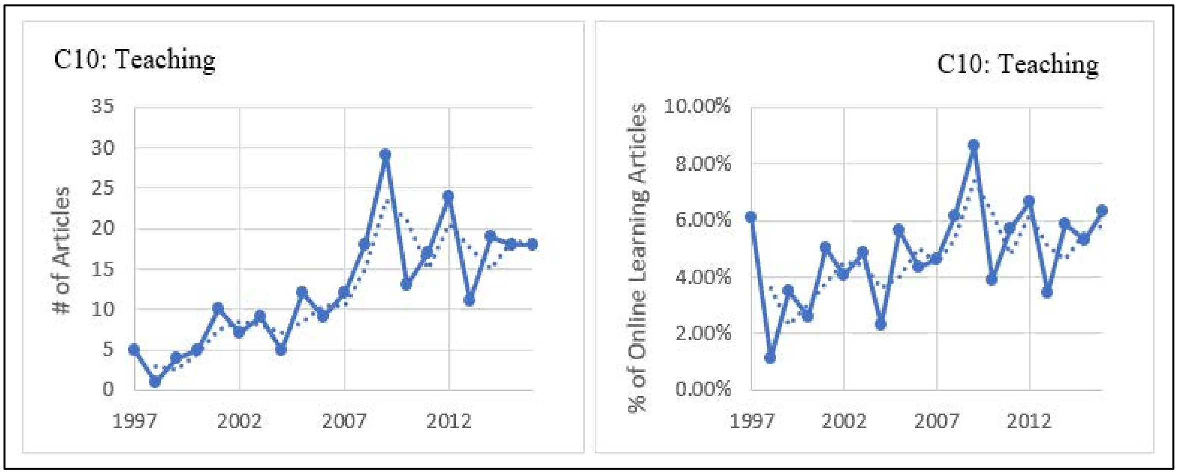

Figure 39. Graphs depicting article distribution in C10: Teaching.

C11: Journal articles. The cluster is comprised of 41 articles focused on information pertaining to journal articles in online learning. This cluster is the smallest in terms of the number of articles and comprises .85 % of all online learning articles examined in this study. The most frequent stemmed terms, in order, are *journal*, *articl*, *publish*, *write*, and *paper*. Three of the articles in this cluster are (a) "The Past, Present, and Future of Research in Distance Education: Results of a Content Analysis" (Lee, Driscoll, & Nelson, 2004), (b) "An Analysis of High Impact Scholarship and Publication trends in Blended Learning" (Halverson, Graham, Spring, & Drysdale, 2012), and (c) "Learning How to Write Effectively for Academic Journals: A Case Study Investigating the Design and Development of a Genre-Based Writing Tutorial System" (Lo, Liu, & Wang, 2014). The average publication year is 2009, and the median year is 2011. The most articles, six, were published in 2013. During the six-year period of 1998-2002, a total of three articles were published. This article publication trend, shown in Figure 40, fluctuated but demonstrated an overall increase after 2002 until 2013. However, the years 2015 and 2016 experienced a steep decline compared to the prior 10 years.

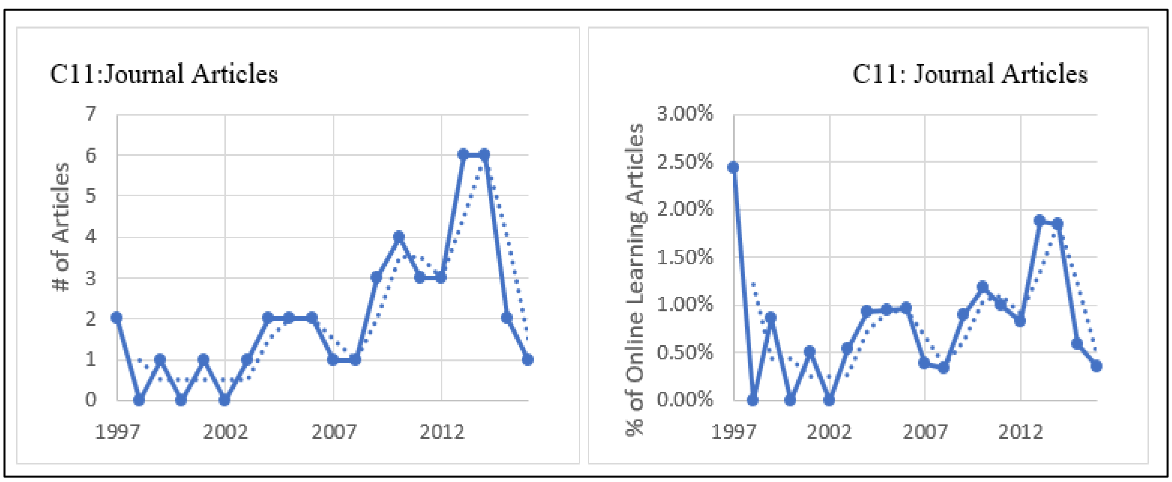

Figure 40. Graphs depicting article distribution in C11: Journal Articles.

C12: Multimedia. This cluster houses 50 articles focused on a variety of issues pertaining to multimedia in online learning. Slightly over one percent (1.04%) of online learning articles were assigned to this cluster. The most used terms include *video, feedback, digit, lectur,* and *instruct*. Samples of writings for this cluster include (a) “The Effect of Context-Based Video Instruction on Learning and Motivation in Online Courses” (Choi & Johnson, 2005), (b) “The Influence of Asynchronous Video Communication on Learner Social Presence: A Narrative Analysis of Four Cases” (Borup, West, & Graham, 2013), and (c) “A Multimedia-Rich Platform to Enhance Student Engagement and Learning in an Online Environment” (Bledsoe & Simmerok, 2013). The mean publication year stands at 2009, and the median is shared between 2011 and 2012. The most prolific years are 2014 and 2015, with seven articles published each year. However, the greatest percentage among online learning article publications is 1998 at 3.5%. The plotted trend lines depicted in Figure 41 for yearly publications oscillate during the twenty-year period.

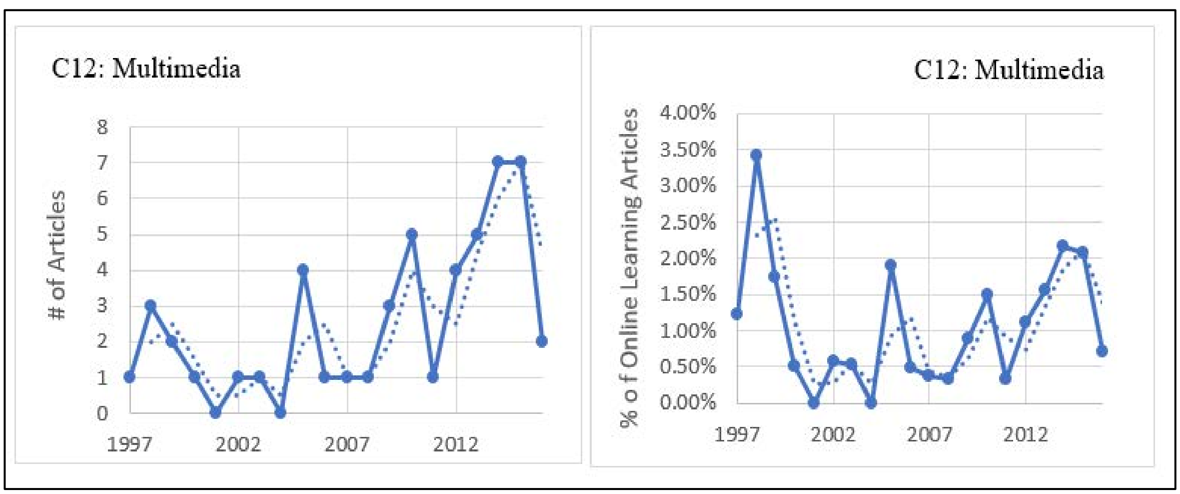

Figure 41. Graphs depicting article distribution in C12: Multimedia.

Increasing Clusters

These three clusters demonstrated an general increase in percentage of artilcle publications during the 20 years. The weighted terms for the three clusters are found in Table 17. A discussion about each cluster is also offered.

Table 17

Centroid Table with Weighted Attribute Values for Increasing Clusters (C13 – C15)

C13 Student Assessment		C14 Learner Motivation and Efficacy		C15 Social Presence	
student	0.042	self	0.199	social	0.202
assess	0.037	efficaci	0.118	presenc	0.195
cours	0.031	regul	0.089	network	0.071
onlin	0.025	motiv	0.075	commun	0.052
test	0.025	learner	0.047	cognit	0.045
feedback	0.024	strategi	0.043	onlin	0.041
perform	0.023	cours	0.036	teach	0.034
effect	0.021	student	0.035	satisfact	0.032
system	0.021	perceiv	0.034	interact	0.032
peer	0.021	learn	0.032	inquiri	0.031
		satisfact	0.032	percept	0.031
		onlin	0.030	instructor	0.03
		srl	0.029	perceiv	0.029
		perform	0.027	student	0.029
		academ	0.027	sens	0.027
		assess	0.026	cours	0.026
		achiev	0.026	relationship	0.025
		internet	0.026	learn	0.024
		web	0.024	particip	0.023
		factor	0.024	media	0.023
		predict	0.023		
		particip	0.023		
		signific	0.022		
		effect	0.022		
		model	0.021		
		measur	0.021		
		predictor	0.021		
		base	0.021		
		teacher	0.021		

C13: Student assessment. Making up 18.67% of all online learning articles in this study, this cluster contains the second largest article collection at 900. Most frequently used terms in this cluster include *student, asses, cours, onlin, test,* and *feedback.* Example articles include (a) "Using JavaScript to Simulate Formative Assessment Questioning in Web-Based Open Learning Materials" (Bowerman, Mansfield, & Sewell, 1997), (b) "Alternative Assessment Approaches for Online Learning Environments in Higher Education" (Reeves, 2000), and (c) "Technology Enhanced Formative Assessment for 21st Century Learning" (Spector, et al., 2016). The average and median publication year is 2010. The highest number of articles published occurred in 2015. As shown in Figure 42, the trend of article publications, both in number and percentage, generally increases throughout the 20-year period.

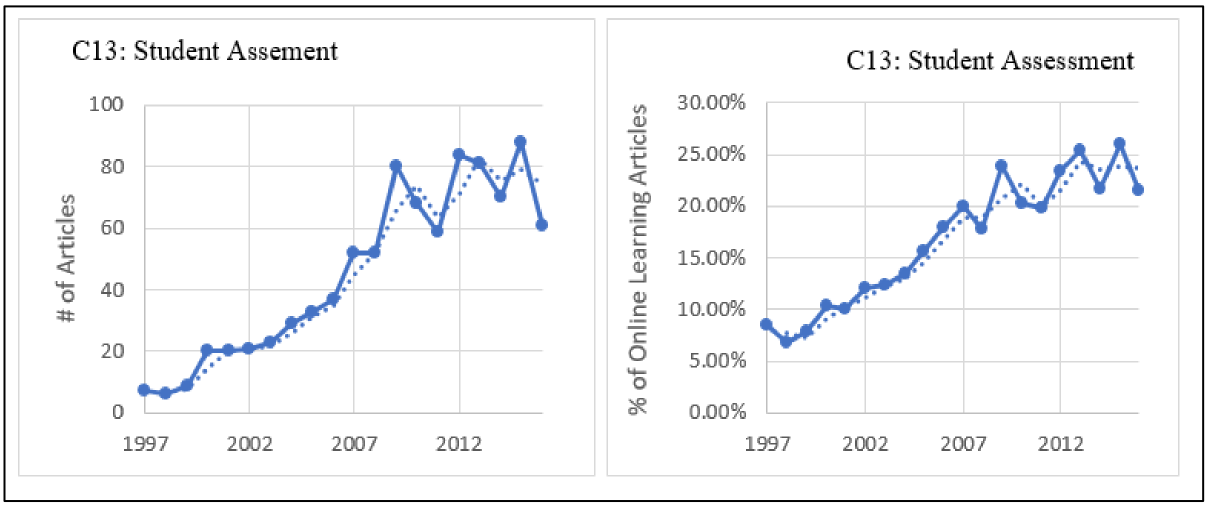

Figure 42. Graphs depicting article distribution in C13: Student Assessment.

C14: Learner motivation and efficacy. This cluster contains 156 articles, representing 3.44% of the selected online learning articles. Most frequently used stemmed terms for this cluster include *self, efficaci, regul, motiv,* and *learner*. Three example articles from this cluster are (a) "Guiding the Independent Learner in Web-Based Training" (J. B. Watson & Rossett, 1999), (b) "The Role of Affective and Motivational Factors in Designing Personalized Learning

Environments" (C. Kim, 2012), and (c) "Expanding Learning Presence to Account for the Direction of Regulative Intent: Self-, Co- and Shared Regulation in Online Learning" (Hayes, Uzuner-Smith, & Shea, 2015). The mean and median year for published articles occurs at the same mark, 2010. However, the most prolific year was 2010, with 18 articles published. The trend lines shown in the graphs of Figure 43 depict a general increase over the 20-year period through 2010, after which fluctuations are noticed, peaking in 2015.

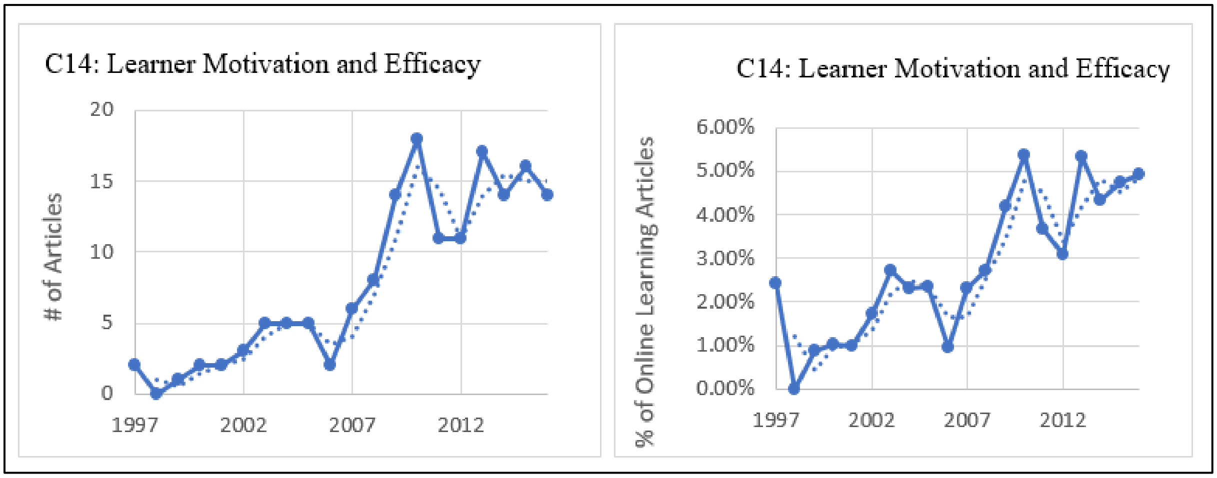

Figure 43. Graphs depicting article distribution in C14: Learner Motivation and Efficacy.

C15: Social presence. This cluster contains 152 articles comprising 3.15% of the online learning articles in the current study. The articles lean towards social presence in online learning settings. The most commonly used stems in this cluster are *social*, *presenc*, *network*, *commun*, and *cognit*. Some sample titles of articles for this cluster are (a) "The Role of Social Comments in Problem-Solving Groups in an Online Class" (Molinari, 2004), (b) "Exploring the Social Competence of Students with Autism Spectrum Conditions in a Collaborative Virtual Learning Environment—The Pilot Study" (Cheng & Ye, 2010), and (c) "Social Presence and Interaction in Learning Environments: The Effect on Student Success" (Kožuh, et al., 2015). The average year of publication is 2010, and the median article occurs in 2011. The last year, 2016, was the most prolific, with 18 articles published. The trend graphs in Figure 44 show a general increase in the number of articles published and in the percentage of online articles, with a noticeable dip for 2015 with the greatest annual increase during the final year, 2016.

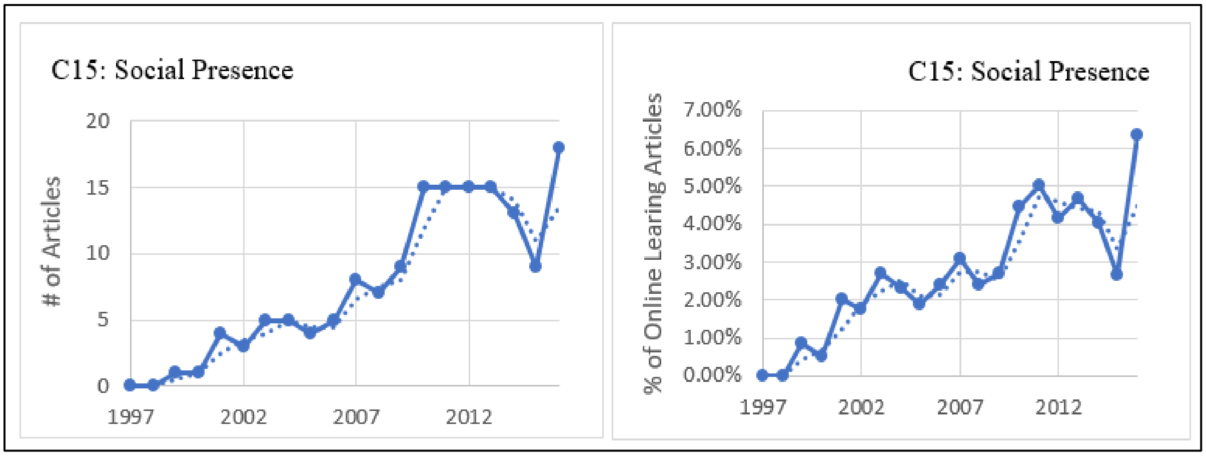

Figure 44. Graphs depicting article distribution in C15: Social Presence

Emerging Clusters

The final three clusters demonstrate a noticeable trending increase during the later portion of the time period. The first several years the publication percentage trending line measured at less than 1%. The weighted terms for this group is located in Table 18. A discussion concerning each cluster is also provided.

C16: Openness and mobility. This cluster contains 105 articles representing 2.18% of the selected online learning articles. The theme of this cluster centers around the concepts of open education, and mobile devices and related issues Three articles included in this grouping are (a) “Mobile Wireless Technology Use and Implementation: Opening a Dialogue on the New Technologies in Education” (S. H. Kim, Holmes, & Mims, 2005), (b) “Opening Up Down Under: The Role of Open Educational Resources in Promoting Social Inclusion in Australia” (Bossu, Bull, & Brown, 2012), and (c) “The Use of a Mobile Learning Management System and Academic Achievement of Online Students” (Han & Shin, 2016). The average publication year is 2010. The following year, 2011, marks the median. In 2012, the next year, more articles were published, 19, than in any other years in this cluster. The general trend, as graphically depicted in Figure 45, shows a general increase through the 20 years. Over the last few years, apparent fluctuations exist.

Table 18

Centroid Table with Weighted Attribute Values for Emerging Clusters (C16 – C18)

C16 Openness and Mobility		C17 Game-Based Learning		C18 MOOCs	
oer	0.17	game	0.414	mooc	0.485
mobil	0.152	plai	0.082	massiv	0.105
devic	0.059	player	0.062	cours	0.067
resourc	0.055	mmorpg	0.058	learner	0.04
metadata	0.039	multiplay	0.044	offer	0.03
textbook	0.036	motiv	0.043	onlin	0.027
educ	0.035	virtual	0.025	particip	0.027
object	0.034	design	0.025	engag	0.025
learn	0.031	solv	0.025	platform	0.022
content	0.029	educ	0.024	coursera	0.022
access	0.029	base	0.024	complet	0.021
wireless	0.024	massiv	0.022	design	0.021
context	0.024	perform	0.021		
technolog	0.023	role	0.021		
applic	0.02	gamer	0.02		

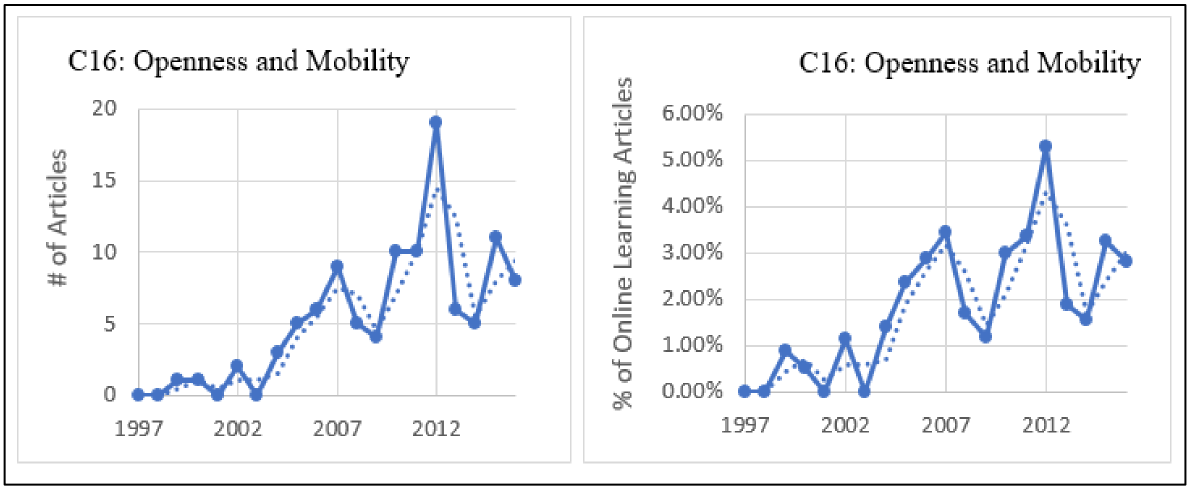

Figure 45. Graphs depicting article distribution in C16: Openness and Mobility.

C17: Game-based learning. Seventy-one articles comprise this cluster, making up 1.47% of the total online learning articles in this analysis. The most frequently used terms occurring in the abstracts are *game, plai, player, mmorpg,* and *multiplay.* The fourth term *mmoopg*, stands for massively multi-player online role-playing game. Example articles from this cluster include (a) "The Design of an Analogical Encoding Tool for Game-Based Virtual Learning Environments" (Williams, Ma, Feist, Richard, & Prejean, 2007), (b) "The Idea Storming Cube: Evaluating the Effects of Using Game and Computer Agent to Support Divergent Thinking" (Huang, Yeh, Li, & Chang, 2010), and (c) "A Solution-Based Intelligent Tutoring System Integrated with an Online Game-Based Formative Assessment: Development and Evaluation" (Hooshyar, et al., 2016). The most articles, 11, published appeared in 2015 with the average and median publication year set as 2012. During the 11 years at the end of the 20-year time frame, 104 of the 105 articles were published. This can be seen on the graphs located in Figure 46. After 2004, the trend in number and percentage of articles generally increase, peaking in 2015. A publication drop within this cluster is noted during the final year, 2016.

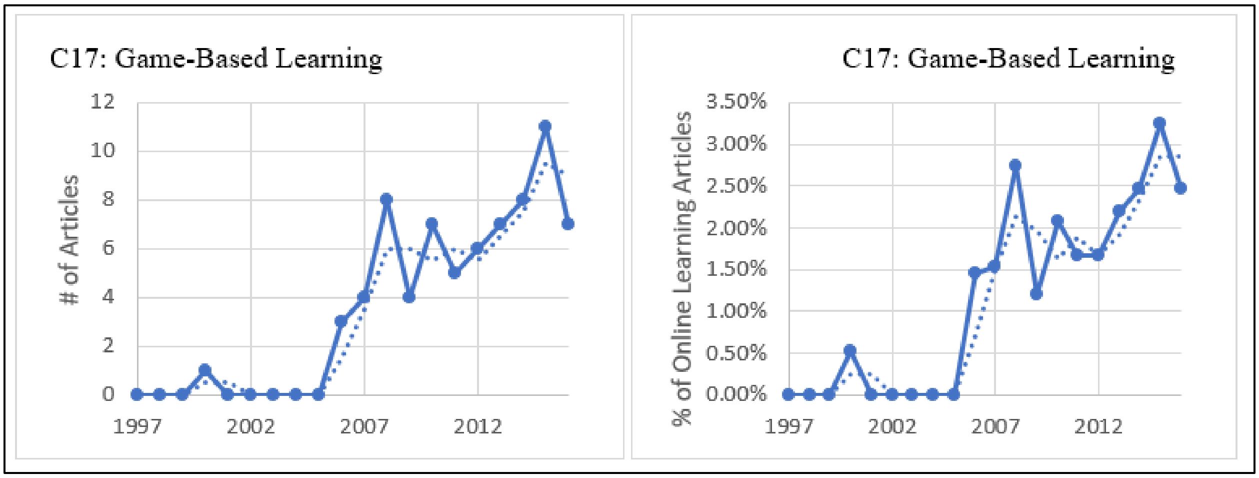

Figure 46. Graphs depicting article distribution in C17: Game-Based Learning.

C18: MOOCs. This cluster centers around issues pertaining to *massive open online courses*, known as MOOCs. Sixty-eight articles comprise the cluster, representing 1.41% of the corpus of documents. The five most commonly occurring terms are *mooc*, *massiv*, *cours*, *learner*, and *offer*. Sample articles include (a) "MOOCs and the AI-Stanford Like Courses: Two Successful and Distinct Course Formats for Massive Open Online Courses" (Rodriguez, 2012), (b) "What Public Media Reveals About MOOCs: A Systematic Analysis of News Reports" (Kovanović V. , Joksimović, Gašević, Siemens, & Hatala, 2015), and (c) "Global Times Call for Global Measures: Investigating Automated Essay Scoring in Linguistically-Diverse MOOCs" (Reilly, et al., 2016). The average and median publication measurement for the MOOC cluster occurs during 2015. For the selected body of online learning literature, the first article in the cluster was published in 2012. As seen with the graphs located in Figure 47, the trend line for both number and percentage of articles is flat at zero and then consistently increases the last few years, with a steep incline noted between 2013 and 2014. The late appearance of this cluster could indicate an potential emerging trend but the steep curve suggesting a possible fad

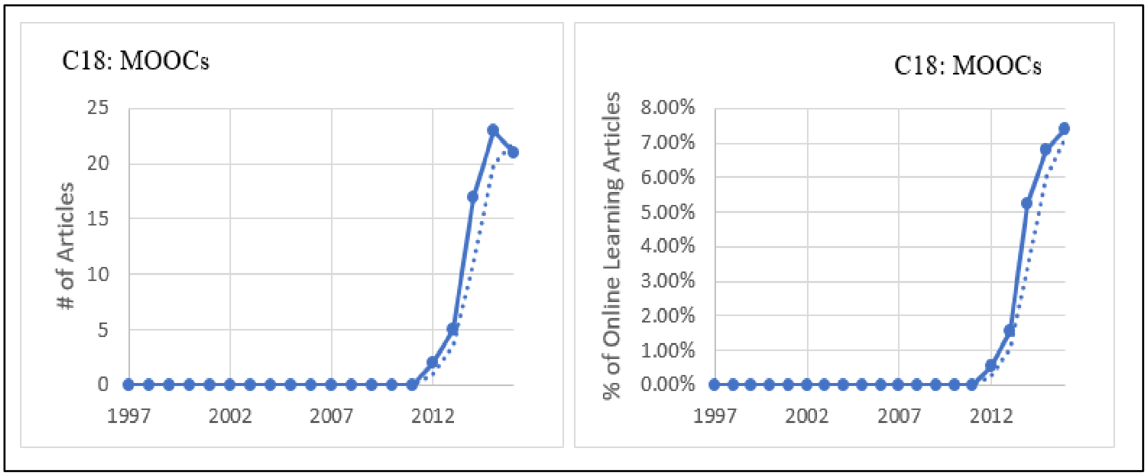

Figure 47. Graphs depicting article distribution in C18: MOOCs.

Summary

The bibliometric analysis of the 4,821 online learning articles revealed journal contribution and yearly publication information. A disparity exists in the contributions from the 15 journals, but this is not surprising. Some of the journals publish considerably more articles per year than others. Also, the aim among the 15 journals differs. So, journals whose emphasis is not technology but more instruction, contributed less than their counterparts steady increase in publications over the 20-year period. The bibliometric analysis focusing on the yearly publication of online learning articles revealed a steady increase over the 20 year period. However, the last few years a possible leveling in both in numbers and percentage of online learning articles published. Though, when looking at online learning article publication numbers in 5-year increments, not including percentages, an increase is apparent between all four time blocks.

The results from the trends in terms exhibited small changes in the frequency of the top three terms. Words containing the stems of *distanc* and *web* declined in overall usage. Though a top ten term early in twenty-year period, their frequency consistently lessened throughout the twenty years. One term, *result,* showed a steady increase in usage, ending in the top ten terms by the last five-year analysis. The top two terms *learn* and *student* remained in two slots during the entire 20 years. In fact, several terms remained in the top ten, but varied within that range.   In word cloud renderings for all time periods, the words *learning* and *student* were prominent, indicating their high frequency use. These trends are also reflected in the clusters, many of which have labels similar to the trending terms.

Over the 20 years, 18 clusters emerged categorizing the articles. Each article is assigned to a cluster. The clusters were identified in the order of the mean year of article publication within the cluster. The identified clusters are:

- C1: Web Issues (2005)
- C2: Distance Education (2006)
- C3 Internet (2006)
- C4: Technology and Education (2008)
- C5: Faculty Concerns (2008)
- C6: Learner Considerations (2008)
- C7: Collaboration (2008)
- C8: Discussion Forums and Communication (2009)
- C9: Virtual Environments and Simulations (2009)
- C10: Teaching (2009)
- C11: Journal Articles (2009)
- C12: Multimedia (2009)
- C13: Student Assessment (2010)
- C14: Learner Motivation and Efficacy (2010)

- C15: Social Presence (2010)
- C16: Openness and Mobility (2010)
- C17: Game-Based Learning (2012)
- C18: MOOCs (2015)

The names of the clusters correspond to their content, with prime consideration being given to the top weighted terms within each cluster. In the individual discussion for each cluster, the top terms and trending information were presented. Clusters that presented a general decline in the number of published articles include C1: Web Issues and C2: Distance Education, and C3: Internet. Clusters demonstrating an overall increase in published articles during the 20-year period are C13: Student Assessment, C14: Learner Motivation and Efficacy, C15: Social Presence, C16: Openness and Mobility, C17: Game-Based Learning, and C18: MOOCs.

A further, more in-depth dialogue of all results is found in Chapter 5. The bibliometrics are discussed and the trends regarding terms and clusters are covered. Future research extensions are presented as well.

CHAPTER 5

DISCUSSION

This study explored the published research in the field of online learning for a recent 20-year period, 1997-2016. The articles originated from 15 journals focused on educational technology (10) and online/distance education (5). Four research questions guided the study:

1. What are the bibliometrics of online learning articles among the 15 journals during the past 20 years (1997-2016)?
2. What are the past, present, and emerging trends of terms regarding online learning within the selected literature during the past 20 years (1997-2016)?
3. What are the thematic clusters of the articles regarding online learning in the 15 journals for the past 20 years (1997-2016)?
4. What are the time trends of the recognized thematic clusters found in the selected articles during the past 20 years (1997-2016)?

The total number of articles meeting the selection criteria numbered 4,821. In order to categorize such a large data set, text-mining techniques were employed to identify the trends and clusters. This chapter reports on the findings and interpretations of the study, offers suggestions for future research, and provides concluding comments.

Findings and Interpretations

Presented in this section is a discussion of the findings as they relate to the bibliometrics and trending terms, generated by the analysis. Discussion of the cluster organization and time trends is also included. Each section offers a unique perspective of the 4,281 articles. The bibliometrics provide an overview of the articles as related to publication years. The trending term and cluster discussions suggest past, present, and emerging trends in vocabulary and topics of study in online learning research

Bibliometric Findings

The bibliometrics of the online learning articles showed an increasing trend in publications during the most of the 20-year period (1997-2016) and a leveling off during the later years. This bore out in both the number of articles published as well as the percentage of articles when compared to the number of articles in all 15 journals. The findings of the bibliometric portion of the study give foundational information regarding the body of articles used in the study.

When examining the contribution percentages among the 15 journals, it was noted that the three with the most articles were from educational technology journals and not from the five field-specific journals. However, this could be due to the fact that those three educational journals contain more articles. The 10 educational technology journals contain 84.17% of all the articles within the 15 journals, and the five more focused journals contain 15.83% of the article used this study. Articles in online learning/distance education journals comprise less than 1/6 of the total articles. However, when considering online learning articles, the journals on that topic contribute 30.12% of the articles, or close to one third of the total. Pie charts graphically displaying percentages are found in Figure 48. The substantial number of contributions from journals outside the specific topic of online learning suggests that general educational technology journals maintain a substantial impact on research regarding online learning.

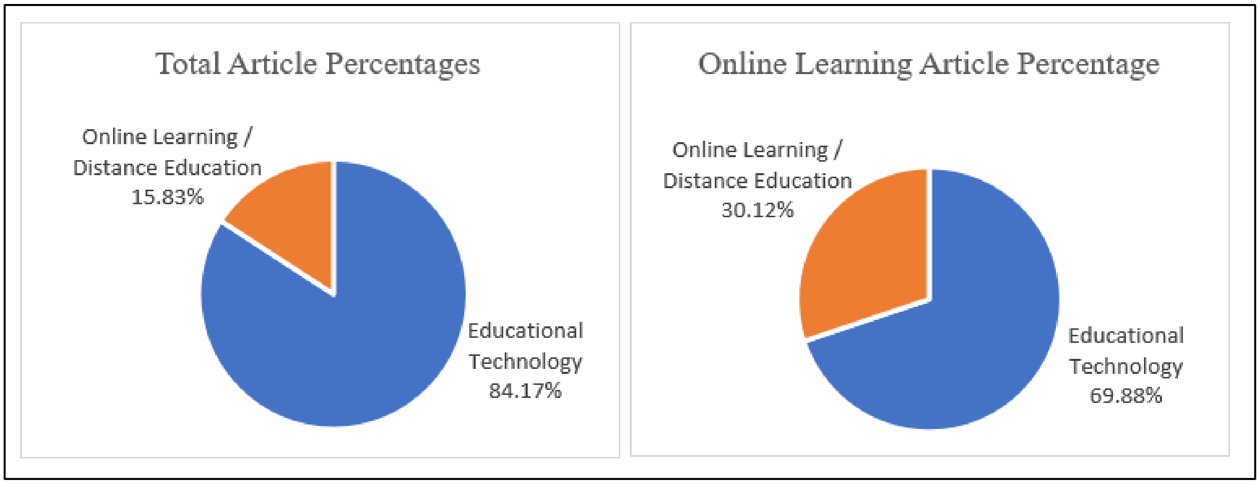

Figure 48. Article distribution according to journal type.

Overall, yearly publication bibliometrics show a general increase during the early years and then a leveling. This observation mimics the bibliometric findings in a trend analysis involving e-learning journal articles (Hung, 2012). This observation could suggest that interest in researching online learning has plateaued over the last several years. Digital network connections within education are becoming more commonplace as many institutions have, and are making substantial investments in, Internet and wireless technologies (Fairlie & Bulman, 2016). An additional explanation is that as these digital connections are becoming stronger and more common, the use of online technology in education becomes an expected commodity and is not necessarily a research focus. Even though the number of published online learning articles may not show an annual increase, when comparing 5-year blocks, an increase is still apparent.

Trends in Terms

In response to the Research Question 2, regarding trends in the terms, the stemmed n-grams for each 5-year period were analyzed. By analyzing ranking stemmed terms within the four small chunks of time, a more comprehensive exploration of the trending terms was possible.

By examining the frequency of terms within each period, snapshots of prominent vocabulary for online learning over specific time periods were developed. Term rankings within each time period were compared and contrasted over the 20-year period to give a broader picture of trending terms within the research. Terms which experienced a marked decline were *distanc* and *web* possibly signaling past trends. An emerging term, *result* was recognized. However, many terms remained relatively stable in their frequency. Among the notable were *learn* and *student* which remained in the top two ranked slots during the entire 20-years period. A chart compiling the top 10 terms with their rankings from each time period is given in Table 19.

Table 19

Top 10 Stemmed N-Gram Rankings

N-gram	1997-2001	2002-2006	2007-2011	2012-2016
learn	1	1	1	2
student	2	2	2	1
educ	3	4	4	4
cours	4	5	6	5
distanc	5	15	26	31
web	6	8	14	34
base	7	6	5	9
technolog	8	10	9	10
develop	9	7	7	6
onlin	10	3	3	3
design	11	9	11	7
teacher	25	20	8	13
learner	26	14	10	14
result	30	26	12	8

Past trends. The terms that experienced a noticeable decline were *web* and *distanc*. Reasons for the decreased use of the term *web* could be that it has become so prevalent that it is assumed present in conversations about online learning. Another possibility is that the abbreviation that is commonly used to refer to the World Wide Web (www) is used instead of the triple term. The term *distanc,* though part of search strings, experienced a sharp decline in use. This supports the notion that although the relationship between distance education and online education are acknowledged (J. L. Moore et al., 2011), online education has progressed at a different pace due to advances in technology and pedagogy (Joksimović, et al., 2015). In fact, according to a keyword calculation on journal articles indexed by SCOPUS, other terms such as *e-learning* and *online learning* consistently increased and surpassed the use of distance education and distance learning around the year 2005 (Joksimović, et al., 2015). The trends of the terms *web* and *distan* mirror clusters with similar labels.

Present trends. The top two stemmed n-grams of *learn* and *student* remained constant throughout the 20-year span. The term *learn* was the most frequently occurring stem during the first three of the four time periods. This corresponds with a prior study of distance education research from 2008 to 2013 where the learning was recognized as a top topic (Bozkurt, et al., 2015). During the last block of time, *learn* dropped to second and *student,* which was the second most used term for the three time periods, took the place of *learn*. Interesting to note is that top terms center around student learning. In fact, the word clouds presented consistently have these two words, *student* and *learning,* as the largest font. It should be noted that *learning* was part of the search string, *distance learning,* but was not used as a singular string. The term *onlin* is the third most occurring term in the last 3 time-periods. However, this could be due to the fact that one of the search strings was *online.*

Other top used terms that remained in the top 10 during the 20 years studied are *educ, cours, develop,* and *technolog*. The stem *educ* consistently remained as the third or fourth term most occurring among the stems. This is not surprising because all the journals included in the study have a focus in education. Also, even though the word *course* was not a single search string, one was *distance education*, which may influence the term count. The term *cours* consistently placed within the fourth and sixth most used stem among the online learning article abstracts. This could be because many of the educational offerings online have been in the form of coursework, and more students are enrolling in online courses (Smith, 2015). Though the word *course* was not part of a search string, *MOOC* was a search string. The *C* in MOOC stands for course and could have influenced the term count, especially in the later years, due to the cluster trends discussed further in this chapter. The term *develop* consistently appears as a top 10 term. Its ranking range is 6-9. Though this term may not be related to online learning on the surface, this term can be used in a variety of related contexts, such as (a) course development, (b) technology development, (c) development of skills, and so forth. Lastly, the regularly appearing stem, *technolog*, ranked between eight and 10 during the 20-year publication history of online learning articles. This abundant use of the stem could be because the delivery of online instruction depends on technology. The topic of emerging technology was noted within e-learning research (Hung, 2012) and in distance education (Bozkurt, et al., 2015). The role technology plays is evident in both the digital devices needed to display content and the distribution mechanisms, wired and wireless, to transfer information.

Emerging term. The term that saw the greatest increase in usage is *result.* In the first 5-year period, it was the thirtieth most occurring term and moved up in ranking for each time period. During the final year span, it was the eighth most used stem. The frequent use of this term could be related to the increased interest in high-stakes testing and outcome-based education. The years in this study are similarly matched with the increased interest in high-stakes testing (Croft, Roberts, & Stenhouse, 2016). Other emerging terms could exist but do not rank in the top 10 terms within the final time period.

Cluster Labels and Time Trends

Research Questions 3 and 4 reference categories within the research and their time trends. Eighteen unique categorical clusters were identified. The initial trends within the clusters seem to mimic the trending terms with the earliest two clusters being C1:Web Issues and C2: Distance Education. Later, possibly emerging clusters were C17: Game-Based Learning and C18: MOOCs. Several of the cluster labels reflect top terms including C6:Learner Considerations and C14: Learning Motivation and Efficacy. Each of the 4,821 articles included in the current study was assigned to a cluster based on the weighted terms contained in each article. An overview of the clusters in terms of size and publication years is displayed in Figure 49.

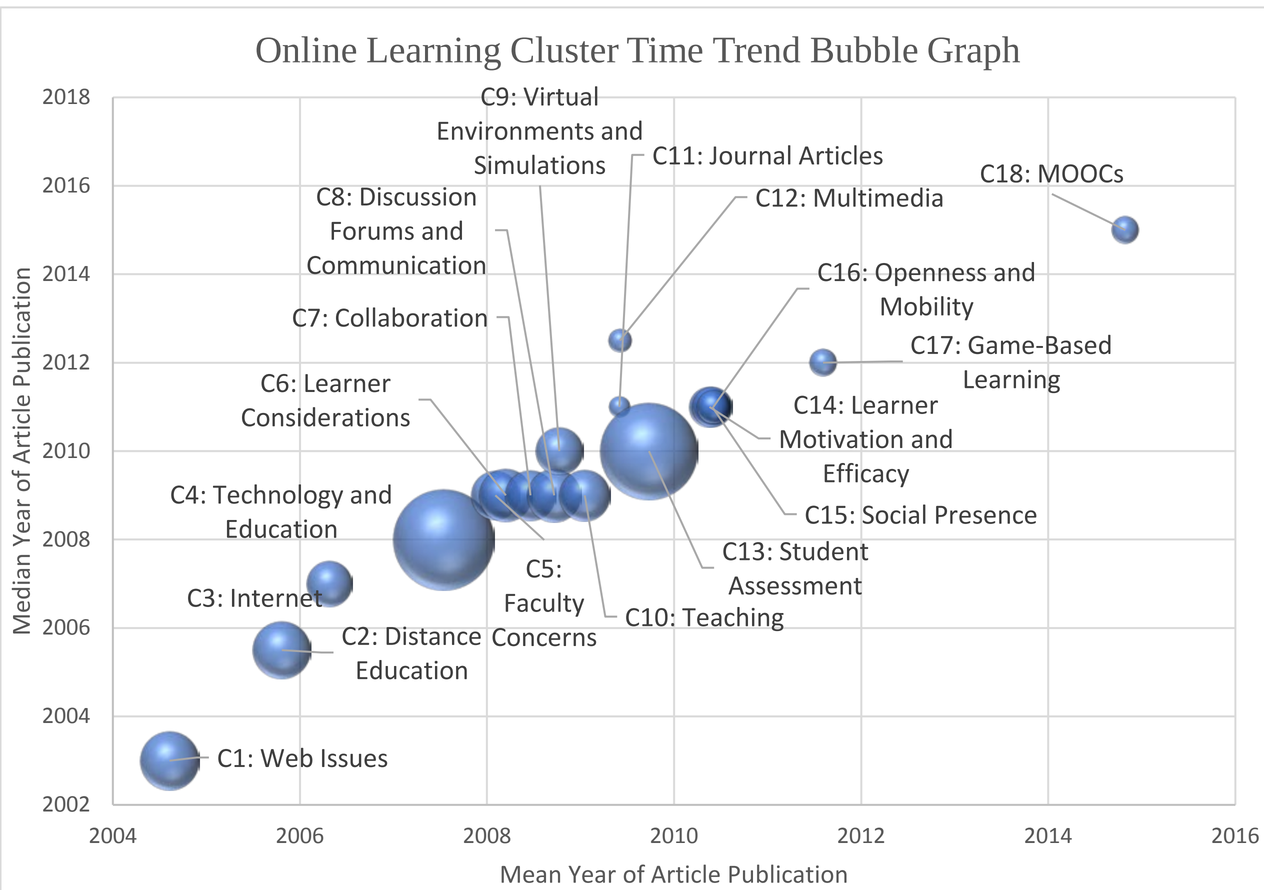

Figure 49. Bubble graph depicting online learning clusters.

In the graph, the horizontal axis represents the average year of publication of the articles contained in the respective clusters. The vertical axis corresponds to the median year. The size of the bubble corresponds to the number of articles in each cluster. For example, cluster C4: Technology and Education is the largest containing 976 articles and is represented by the biggest bubble on the chart. The smallest bubble, C11: Journal Articles, contains the fewest at 41 articles.

Decreasing trends. Two clusters experience a marked decline in article publications during the study period. C1: Web Issues and C2: Distance Education are the two clusters identified as declining. This corresponds with the stemmed terms *web* and *distanc* being identified as past trending terms in the previous section. Though Web 2.0 was recognized as a trend in distance education research for the years 2009-2013 (Bozkurt, et al., 2015). The decline in Web issues could be due to changes and expansion in vocabulary related to internet technology over time.

The prominence of online learning in distance education was recognized in prior literature (Joksimović, et al., 2015; Taylor, 2001). In a sense of reversal, online learning was recognized as a theme in a report on the history of the journal *Distance Equation* (Zawacki-Richter & Naidu, 2016). Other studies of online learning recognized distance education as prominent categories (Bozkurt, et al., 2015; Zawacki-Richter, et al., 2009). Two of the journals connate the term in their name, *Distance Education* and *American Journal of Distance Education.* At one time, online learning could be considered a subset of distance education. However, as technology advances, online learning has become a topic in its own right evolving in a different direction. This notion is reflected in the naming of one of the journals used in this study. One of the journals used in this study, *Online Learning*, changed its name from *Journal of*

Asynchronous Learning Networks in 2014 to better reflect the sponsoring organization's current mission (Online Learning Consortium, 2017). Online learning still continues to have a role in distance education, but its use has grown to maintain a broader application.

A third cluster C3: Internet remained fairly stable in publication numbers throughout the 20 years. However, a noted decline as percentage of online learning articles is evident. The decline was not as steep as the previous two clusters. This behavior corresponds with the trend of the stem *internet* within the literature. During the first five-year period, 1997-2001, the stem was ranked 12th among frequently used terminology within the literature. For the subsequent three five-year periods, the use of the term, did not rank above 25. The percentage decline in *internet* could be explained by a widening vocabulary towards aspects in online learning or perhaps a shift in wireless and mobile technology as the word Internet can have a more hardware connotation. A recent study published in 2015 recognized internet aspects of online learning as a top keyword (Bozkurt, et al., 2015). The apparent declining research interest could be attributed to the Internet technology becoming so commonplace that it is not mentioned or acknowledged as much as it once was.

Steady trends. C4: Technology and Education, and C5: Faculty Concerns showed fluctuations, but overall remained relatively stable from 1997-2016 for article publication numbers. Each of these clusters has a technology implication that has been, and will likely continue to be, used in education. This could explain the stability seen in the clusters. That same study also separated the current cluster C4 into two separate groups, technology and educational technology (Bozkurt, et al., 2015). In the present research, cluster C4: Technology and Education is the most populated cluster, giving further evidence for its continued research interest now and in the near future. Also, most articles in the study hailed from recognized educational technology journals. The cluster C5: Faculty Concerns was also as a category in a prior study (Zawacki-Richter, et al., 2009). Faculty concerns can encompass professional development and continuing education both of which are popular topics of research in general.

Peaked trends. The next three clusters, C6: Learner Considerations, C7: Collaboration, and C8: Discussion Forums and Communication, possess similar trends. A noted increase trend in the middle or end of the time period, followed by a decline is apparent. These tendencies indicate a decline in research that focused on learner-centered instruction or collaboration as they related to online learning. Categories of research similar to C6: Learner Considerations focusing on student support and perception appeared in prior research (Bozkurt, et al., 2015; Zawacki-Richter, et al., 2009). Collaboration was also identified as a category of distance education (Bozkurt, et al., 2015; Zawacki-Richter & Naidu, 2016) The topic of discussion forums and communication has been a popular category recognized in at least three prior studies (Bozkurt, et al., 2015; Zawacki-Richter, et al., 2009, Zawacki-Richter & Naidu, 2016)).

Fluctuating trends. The subsequent four clusters — C9: Virtual Environments and Simulations, C10: Teaching, C11: Journal Articles, and C12: Multimedia— demonstrate varied movements within two-decade window. In these clusters, no consistent trend was noted. This could be due to possible multiple special issues on a specific topic or that research interests in these topics generally fluctuate over time. These categories have also been identified in prior literature regarding online learning. Closely related to C9: Virtual Environments and Simulations, Bozkurt, et al. (2009) noted virtual school as topic, and Hung (2012) named simulation as a cluster in the broader category of e-learning. A classification area in distance education involving research methods in online learning was presented (Bozkurt, et al., 2015) which is associated with C10: Journal Articles. Multimedia was another research cluster named by Hung (2012) for e-learning.

Increasing trends. Several clusters showed an increase in published articles over the 20-year period. This is not surprising since the number of published articles in online learning has generally increased over the time period, as noted in the discussion of the bibliometric analysis. Three clusters demonstrated the increasing trending throughout the 20-year time frame. These three clusters include C13: Student Assessment, C14: Learner Motivation and Efficacy, and C15: Social Presence. The first two of these clusters demonstrated a steady increase with publication fluctuations noted during the last few years. These clusters have a student-centered component and similar publication trends within the 20-year period. One study reported motivation as a category of research (Bozkurt, et al., 2015) which is incorporated into C14: Learner Motivation and Efficacy.

In recent years, social media has become commonplace in daily life and is making its mark in the education arena. The cluster C15: Social Presence suggests this use of instructional technology. Though the research trend loosely follows the previous two clusters, a noted sharp increase is evident during the final year. Due to the drop in overall online learning articles during the last year, most clusters do not show an increase that final year. This recent increase could signal an emerging research interest in social media, as a lack of predominant presence in education has been recognized. Though social media use is pervasive and its use in education seems inevitable, few social media strategies have been launched in university settings (Kilis, Gülbahar, & Rapp, 2016). Perhaps this late surge in research in this area could prompt wider implementation of social media strategies in education.

Emerging trends. The final three clusters also demonstrate an increasing trend in publications. Though during the first several years, the percentage of publications in online learning remained at or below 1%, causing them to be considered emerging. Consider cluster C16: Openness and Mobility. When looking specifically at the data, it coincides with Hwang's and Tsai's (2011) observation of an uptick in research in mobile learning from 2008 through 2010. Interesting to note that article publications since 2010 have fluctuated. This could be due to the constantly changing technology in this area, such as smart phones and tablets. As digital-based technology becomes smaller, faster, and less expensive, mobile technology becomes more easily obtainable for and accessible to students.

The cluster C17: Game-Based Learning started slowly in publications but showed substantial increases since 2005. This corresponds with another study indicating significant increases in digital game-based learning from 2001 to 2010 (Hwang & Wu, 2012). The increase aligns with the growing number of researchers developing games to promote 21st century learning skills among students (Quain & Clark, 2016). Gaming in education has been, and will likely continue to be, an interest to researchers and educators.

The final cluster, C18: MOOCs, appears to be an outlier on the bubble graph in Figure 49. This cluster has experienced recent phenomenal growth in published articles suggesting that MOOCs could be an emerging trend in online learning research. This is in agreement with a recent study that claims that the extent of MOOC research is likely to increase in the coming years (Bozkurt, Akgün-Özbek, & Zawacki-Richter, 2017). Future research will tell if the MOOC trend in education research continues past 2016.

Summary of Findings

During the course of this research, several themes were recognized. Many of the results confirm prior research in the field, and some newer discoveries were revealed. Major findings from this research encompass all four areas of analysis.

- The relationship between online learning and distance education is evident throughout the research. The literature review showed that early in the research period online learning could be considered a subset of distance education, but online learning is, thanks to technology, moving in its own direction. Within the trending of terms and cluster analysis, distance references were more frequently used in the early years, with a consistent decline noted through the 20-year period.

- The focus on learning is apparent throughout the literature based on the trending terms and cluster analysis. The emphasis on learning and students is evident and easily noticed in the word clouds of the article abstracts throughout the entire 20 year history of the literature.

- The connection of educational technology research and online learning research was revealed through the bibliometric and cluster analysis. The publication yearly variations of the online learning and the other educational technology articles were similar. Also, the largest recognized cluster category of online learning was educational technology.

- Recognized emerging themes include (a) openness and mobility, (b) game-based learning, and (c) MOOCs. These three maintained the latest publication dates and are recognized as potential areas to receive more attention in future research.

Recommendations for Future Research

The findings presented set a foundation for future research in online learning. This study focused on one identified category from a larger study of educational technology research (Natividad, 2016). Future work could take this effort a step further by providing additional examination into one of the identified categorical clusters from this overview of online learning research. The summaries presented could provide foundational information towards research focused on one of the identified topics.

A possible extension of research would be to follow one linguistic term, for example *e-learning,* throughout the research literature over an extended period of time. Investigating the definitions and evolving applications of educational technology terms would provide insight into appropriate current and future use. With the rapidly changing technology landscape, this type of research could be a beacon for standardization and clarification of terminology.

Likewise, a study could occur specifically examining the relationship of online learning and distance education. While revealed as a theme throughout this study, further exploration of the connection could provide valuable information to both.   Similarities and differences in their respective research could be analyzed. Also, comparing and contrasting how each topic is covered in educational technology journals could be useful for researchers and practitioners.

An extension suggestion to the research involves further investigation of each of the clusters. For example, research could be carried out to examine the country of origins for articles contained in each cluster. This could address the intercultural aspects of distance learning which was noted as a neglected area in prior research (Zawacki-Richter, et al., 2009). Also, a more detailed study of cluster size could be carried out to further understand any implications. Additionally, a comparison of individual cluster trends with the Gartner Hype Cycle (Gartner, Inc., 2017) could provide insight regarding the adoption and maturity of related learning technology applications. By further examining the clusters, a deeper understanding of the topics and trends could transpire.

The online learning articles selected for the study were based on the results of several searches using specific search strings. An interesting modification could be using a different list of text strings for searching the articles from the 15 journals. Using previous recognized terms within the research, such as the 10 shown in Figure 2. While derivative of the first six terms were used, the final four, (a) distributed learning, (b) computer aided learning, (c) computer assisted learning, and (d) computer mediated learning (Joksimović, et al., 2015), were not part of a search string. These terms could be added to the existing search strings or the 10 terms could be used without influence from other terminology.

Another suggestion is to use the same data but a different methodology, then compare and contrast the results. One possibility would be to employ a different process for grouping the research. There are other categorization algorithms that could be used for a comparison study, such as hierarchical clustering or factor analysis, that use different mathematical calculations to organize data. Comparisons between the current results and results from using a different grouping mechanism could provide further insight in to the status of online learning research.

Conclusion

Spector, Johnson, and Young (2014) noted that there is tremendous value in using a variety of methods in the exploration of foundational areas within educational technology. This study used text mining techniques to discover trends within published research of online learning across multiple journals. The automation of the analysis allowed for a larger article set than is feasible in manual analysis. The work extended Natividad's (2016) research through further examination the identified category of online learning within educational technology peer-reviewed published research. Fifteen highly regarded academic journals served as the data source from which past, present and emerging trends within the online learning research were identified and discussed.

This study served to fill an identified gap of online learning research analysis across multiple journals. Results from this metatrend analysis generally align with and extend prior meta-efforts in online learning. These findings are a good source for future researchers planning to contribute to the research of online learning or in one of the identified cluster categories. Additionally, researchers and practitioners can utilize information concerning trends in online learning leading to future projects such as special issues or conference sessions. Examining past research through advanced analytic techniques, as done in this study, informs educators and instructional designers about trends within online learning. Identified trends can provide leverage to educational and professional institutions that may enhance the design, development, and delivery of online instruction.

APPENDIX A

IRB

IRB Number:	17-307
Title of study:	Examining Trends in Online Learning: A Meta-Trend Analysis using Text Mining Reviewing 20 years of Research Literature
Supervising Investigator:	Lin Lin
Supervising Investigator e-mail:	[redacted]
Student Investigator:	Heather Keahey
Student Investigator e-mail:	[redacted]

Hello Dr. Lin,

The UNT Institutional Review Board has jurisdiction to review proposed "research" with "human subjects" as those terms are defined in the federal IRB regulations. The regulations define research as "a systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge." The phrase "human subjects" is defined as "a living individual *about* whom an investigator (whether professional or student) conducting research obtains (1) data through intervention or interaction with the individual, or (2) identifiable private information.

Based on the description of your proposed study in your IRB Application, your study will not involve "human subject research" therefore review and approval by the UNT IRB is not needed.

We appreciate your efforts, however, to comply with the federal regulations and sincerely thank you for your IRB application submission!

Thank You,

Jillian Byrne-Sweeney, MA
Research Analyst II, Research Integrity and Compliance
Office of Research and Innovation
University of North Texas
[redacted]
Website: Research Integrity and Compliance

APPENDIX B

ADDITIONAL JOURNAL HISTORIES

Publication Title	Publication History as of May 2017
The American Journal of Distance Education (Selected)	30 years – since 1987
Distance Education (Selected)	37 years – since 1980
E-Learning and Digital Media	13 years – since 2004
E-Learning and Education	12 years – since 2005
Electronic Journal of e-Learning	14 years- since 2003
European Journal of Open, Distance, and E-Learning	13 years -
International Journal of Distance Education Technologies	14 years – since 2003
International Journal of E-Learning	15 years – since 2002
*International Journal E-Learning & Distance Education**	31 years – since 1986
International Journal of Instructional Technology and Distance Learning	13 years – since 2004
International Journal of Online Pedagogy and Course Design	7 years – since 2011
International Review of Research in Open and Distance Learning	16 years – since 2000
Journal of Interactive Online Learning	14 years – since 2003
Online Journal of Distance Learning Administration	19 years – since 1998
Online Learning (Selected)	20 years – since 1997
Open Learning: The Journal of Open, Distance and e-Learning (Selected)	31 years – since 1986
Quarterly Review of Distance Education	17 years – since 2000

*Not selected due to lack of reliable impact factor information

APPENDIX C

LIST OF ENGLISH AND RESEARCH STOP WORDS

a
about
above
across
after
again
against
aim
all
almost
alone
along
already
also
although
always
among
an
and
another
any
anybody
anyone
anything
anywhere
are
area
areas
around
article
as
ask
asked
asking
asks
at
author
authors
away
b
back
backed
backing
backs
be
became
because
become
becomes
been
before
began
behind
being
beings
best
better
between
big
both
but
by
c
came
can
cannot
case
cases
certain
certainly
clear
clearly
come
could
d
describes
did
differ
different
differently
discuss
discusses
do
does
done
down
down
downed
downing
downs
during
e
each
early
either
end
ended
ending
ends
enough
even
evenly
ever
every
everybody
everyone
everything
everywhere
examine
examined
examines
explore
explores
f
face
faces
fact
facts
far
felt
few
find
finds
first
focus
for
four
from
full
fully
further
furthered
furthering
furthers
g
gave
general
generally
get
gets
give
given
gives
go
goal
going
good
goods
got
great
greater
greatest
group
grouped
grouping
groups
h
had
has
have
having
he
her
here
herself
high
high
high
higher
highest
him
himself
his
how
however
if
important
in
interest
interested
interesting
interests
into
investigate
investigated
investigates
is
it
its
itself
j
just
k
keep
keeps
kind
knew
know
known
knows
l
large
largely
last
later
latest
least
less
let
lets
like
likely
long
longer
longest
m
made
make
making
man
many
may
me
member
members
men
might
more
most
mostly
mr
mrs
much
must
my
myself
n
necessary
need
needed
needing
needs
never
new

newer
newest
next
no
nobody
non
noone
not
nothing
now
nowhere
number
numbers
o
of
off
often
old
older
oldest
on
once
one
only
open
opened
opening
opens
or
order
ordered
ordering
orders
other
others
our
out
over
p
paper
part
parted
parting
parts
per
perhaps
place
places
point
pointed
pointing
points
possible
present
present
presented
presenting
presents
presents
problem
problems
purpose
put
puts
q
quite
r
rather
really
recent
recently
reports
research
right
right
room
rooms
s
said
same
saw
say
says
second
seconds
see
seem
seemed
seeming
seems
sees
several
shall
she
should
show
showed
showing
shows
side
sides
since
small
smaller
smallest
so
some
somebody
someone
something
somewhere
state
states
still
study
such
sure
t
take
taken
than
that
the
their
them
then
there
therefore
these
they
thing
things
think
thinks
this
those
though
thought
thoughts
three
through
thus
to
today
together
too
took
toward
turn
turned
turning
turns
two
u
under
until
up
upon
us
use
used
uses
v
very
w
want
wanted
wanting
wants
was
way
ways
we
well
wells
went
were
what
when
where
whether
which
while
who
whole
whose
why
will
with
within
without
work
worked
working
works
would
writer
writers
x
y
year
years
yet
you
young
younger
youngest
your
yours
z

APPENDIX D

TOP 100 N-GRAMS WITH TF-IDF CALCULATIONS

Top 100 n-grams for online learning articles 1997-2001

	N_Gram	Total	per Article
1	learn	1046	1.54
2	student	945	1.39
3	educ	885	1.30
4	cours	641	0.94
5	distanc	523	0.77
6	web	504	0.74
7	base	493	0.73
8	technolog	453	0.67
9	develop	444	0.65
10	onlin	410	0.60
11	design	371	0.55
12	internet	348	0.51
13	commun	345	0.51
14	environ	309	0.46
15	comput	301	0.44
16	system	296	0.44
17	interact	287	0.42
18	instruct	281	0.41
19	distance_educ	280	0.41
20	teach	274	0.40
21	univers	256	0.38
22	inform	237	0.35
23	discuss	234	0.34
24	support	228	0.34
25	teacher	224	0.33
26	learner	218	0.32
27	experi	205	0.30
28	collabor	194	0.29
29	model	191	0.28
30	result	191	0.28
31	web_bas	187	0.28
32	process	182	0.27
33	evalu	179	0.26
34	project	177	0.26
35	effect	169	0.25
36	program	165	0.24
37	provid	160	0.24
38	school	160	0.24
39	particip	159	0.23
40	issu	156	0.23
41	train	152	0.22
42	includ	151	0.22
43	tool	146	0.22
44	institut	144	0.21
45	distance_learn	142	0.21
46	line	140	0.21
47	activ	138	0.20
48	learning_environ	134	0.20
49	virtual	134	0.20
50	approach	133	0.20
51	data	133	0.20
52	level	132	0.19
53	time	130	0.19
54	resourc	129	0.19
55	network	128	0.19
56	offer	128	0.19
57	faculti	127	0.19
58	knowledg	127	0.19
59	relat	127	0.19
60	classroom	125	0.18
61	practic	123	0.18
62	analysi	119	0.18
63	cultur	118	0.17
64	facilit	118	0.17
65	implement	113	0.17
66	tradit	113	0.17
67	access	112	0.16
68	world	112	0.16
69	content	111	0.16
70	materi	110	0.16
71	wide	109	0.16
72	instructor	108	0.16
73	deliveri	106	0.16
74	site	105	0.15
75	method	103	0.15
76	suggest	102	0.15
77	electron	101	0.15
78	requir	96	0.14
79	assess	94	0.14
80	question	93	0.14
81	academ	90	0.13
82	factor	90	0.13
83	studi	90	0.13
84	chang	89	0.13

Top 100 n-grams for online learning articles 1997-2001 (continued)

85	creat	89	0.13
86	integr	89	0.13
87	strategi	87	0.13
88	survei	87	0.13
89	asynchron	86	0.13
90	set	86	0.13
91	address	84	0.12
92	class	84	0.12
93	manag	84	0.12
94	role	83	0.12
95	social	83	0.12
96	applic	82	0.12
97	individu	82	0.12
98	qualiti	81	0.12
99	find	80	0.12
100	skill	79	0.12

Top 100 n-grams for online learning articles 2002-2006

	N_Gram	Total	per Article
1	learn	2382	2.40
2	student	1839	1.85
3	onlin	1385	1.39
4	educ	983	0.99
5	cours	973	0.98
6	base	781	0.79
7	develop	696	0.70
8	web	647	0.65
9	design	597	0.60
10	technolog	542	0.55
11	environ	496	0.50
12	discuss	494	0.50
13	commun	484	0.49
14	learner	475	0.48
15	distanc	471	0.47
16	support	434	0.44
17	instruct	426	0.43
18	interact	420	0.42
19	system	405	0.41
20	teacher	391	0.39
21	collabor	389	0.39
22	effect	381	0.38
23	model	372	0.37
24	univers	364	0.37
25	teach	358	0.36
26	result	356	0.36
27	process	353	0.36
28	particip	331	0.33
29	experi	319	0.32
30	inform	318	0.32
31	web_bas	305	0.31
32	evalu	304	0.31
33	knowledg	303	0.31
34	comput	294	0.30
35	internet	294	0.30
36	tool	290	0.29
37	program	265	0.27
38	learning_environ	263	0.26
39	practic	260	0.26
40	assess	256	0.26
41	school	256	0.26
42	approach	255	0.26
43	online_learn	255	0.26
44	activ	252	0.25
45	provid	239	0.24
46	includ	238	0.24
47	analysi	236	0.24
48	instructor	219	0.22
49	content	218	0.22
50	distance_educ	218	0.22
51	time	210	0.21
52	social	207	0.21
53	level	206	0.21
54	studi	201	0.20
55	project	195	0.20
56	suggest	193	0.19
57	resourc	190	0.19
58	role	189	0.19
59	skill	189	0.19
60	issu	187	0.19
61	class	184	0.19
62	data	184	0.19
63	find	180	0.18
64	faculti	175	0.18
65	virtual	171	0.17
66	factor	170	0.17
67	structur	170	0.17
68	self	168	0.17
69	qualiti	166	0.17
70	perform	163	0.16
71	relat	162	0.16
72	distance_learn	158	0.16
73	classroom	157	0.16
74	institut	157	0.16
75	offer	157	0.16
76	strategi	157	0.16
77	context	155	0.16
78	facilit	154	0.16
79	train	153	0.15
80	materi	152	0.15
81	question	152	0.15
82	academ	151	0.15
83	integr	151	0.15
84	online_cours	150	0.15

Top 100 n-grams for online learning articles 2002-2006

85	tradit	150	0.15
86	manag	149	0.15
87	increas	148	0.15
88	based_learn	147	0.15
89	engag	147	0.15
90	implement	147	0.15
91	improv	146	0.15
92	reflect	146	0.15
93	access	144	0.15
94	differ	143	0.14
95	requir	137	0.14
96	help	136	0.14
97	specif	136	0.14
98	task	136	0.14
99	method	134	0.13
100	network	134	0.13

Top 100 n-grams for online learning articles 2007-2011

	N_Gram	Total	per Article
1	learn	4238	2.78
2	student	3332	2.19
3	onlin	2102	1.38
4	educ	1570	1.03
5	base	1103	0.72
6	cours	1087	0.71
7	develop	994	0.65
8	teacher	884	0.58
9	technolog	872	0.57
10	learner	863	0.57
11	design	851	0.56
12	result	800	0.52
13	system	800	0.52
14	web	749	0.49
15	environ	748	0.49
16	effect	729	0.48
17	discuss	722	0.47
18	support	670	0.44
19	interact	664	0.44
20	commun	638	0.42
21	teach	632	0.41
22	experi	607	0.40
23	model	605	0.40
24	particip	603	0.40
25	knowledg	602	0.40
26	distanc	584	0.38
27	inform	576	0.38
28	activ	522	0.34
29	univers	518	0.34
30	social	507	0.33
31	instruct	506	0.33
32	process	489	0.32
33	internet	477	0.31
34	collabor	461	0.30
35	practic	461	0.30
36	virtual	461	0.30
37	assess	449	0.29
38	school	447	0.29
39	data	437	0.29
40	tool	428	0.28
41	analysi	426	0.28
42	self	423	0.28
43	comput	420	0.28
44	evalu	408	0.27
45	provid	406	0.27
46	level	400	0.26
47	find	389	0.26
48	program	385	0.25
49	factor	382	0.25
50	approach	363	0.24
51	learning_environ	363	0.24
52	web_bas	362	0.24
53	content	356	0.23
54	online_learn	346	0.23
55	test	340	0.22
56	perform	337	0.22
57	includ	327	0.21
58	indic	327	0.21
59	signific	324	0.21
60	suggest	324	0.21
61	strategi	317	0.21
62	relat	305	0.20
63	differ	301	0.20
64	method	301	0.20
65	qualiti	298	0.20
66	time	293	0.19
67	identifi	289	0.19
68	understand	287	0.19
69	instructor	283	0.19
70	studi	282	0.19
71	distance_educ	281	0.18
72	context	280	0.18
73	user	279	0.18
74	project	272	0.18
75	skill	269	0.18
76	improv	266	0.17
77	propos	266	0.17
78	peer	260	0.17
79	help	259	0.17
80	perceiv	257	0.17
81	institut	255	0.17
82	cognit	251	0.16
83	motiv	251	0.16

Top 100 n-grams for online learning articles 2007-2011

84	implement	248	0.16
85	structur	248	0.16
86	role	246	0.16
87	train	246	0.16
88	engag	241	0.16
89	integr	241	0.16
90	facilit	240	0.16
91	feedback	240	0.16
92	issu	239	0.16
93	survei	236	0.15
94	posit	235	0.15
95	success	235	0.15
96	found	234	0.15
97	construct	232	0.15
98	impact	231	0.15
99	reflect	231	0.15
100	access	229	0.15

Top 100 n-grams for online learning articles 2012-2016

	N_Gram	Total	Occurences per Article
1	student	4336	2.67
2	learn	4136	2.55
3	onlin	2930	1.80
4	educ	1746	1.07
5	cours	1624	1.00
6	develop	1050	0.65
7	design	1034	0.64
8	result	1005	0.62
9	base	1004	0.62
10	technolog	908	0.56
11	particip	898	0.55
12	effect	882	0.54
13	teacher	844	0.52
14	learner	826	0.51
15	discuss	766	0.47
16	environ	760	0.47
17	interact	691	0.43
18	social	689	0.42
19	teach	688	0.42
20	support	653	0.40
21	univers	630	0.39
22	activ	629	0.39
23	system	624	0.38
24	commun	616	0.38
25	inform	614	0.38
26	experi	613	0.38
27	assess	610	0.38
28	model	610	0.38
29	data	605	0.37
30	perform	596	0.37
31	distanc	528	0.32
32	knowledg	525	0.32
33	analysi	523	0.32
34	web	520	0.32
35	practic	516	0.32
36	self	508	0.31
37	find	506	0.31
38	approach	503	0.31
39	level	501	0.31
40	school	499	0.31
41	instruct	498	0.31
42	collabor	486	0.30
43	tool	464	0.29
44	virtual	464	0.29
45	process	448	0.28
46	evalu	446	0.27
47	program	435	0.27
48	factor	427	0.26
49	online_cours	425	0.26
50	engag	419	0.26
51	indic	416	0.26
52	provid	411	0.25
53	strategi	406	0.25
54	content	400	0.25
55	relat	394	0.24
56	suggest	383	0.24
57	academ	378	0.23
58	test	376	0.23
59	motiv	369	0.23
60	improv	366	0.23
61	online_learn	364	0.22
62	mooc	362	0.22
63	instructor	361	0.22
64	signific	360	0.22
65	peer	358	0.22
66	skill	350	0.22
67	time	350	0.22
68	studi	343	0.21
69	includ	340	0.21
70	resourc	340	0.21
71	survei	340	0.21
72	found	335	0.21
73	learning_environ	335	0.21
74	institut	333	0.20
75	cognit	321	0.20
76	question	317	0.20
77	understand	316	0.19
78	increas	315	0.19
79	posit	311	0.19
80	achiev	310	0.19
81	differ	310	0.19
82	feedback	308	0.19

Top 100 n-grams for online learning articles 2012-2016 (continued)

83	method	305	0.19
84	qualiti	302	0.19
85	implement	295	0.18
86	impact	289	0.18
87	relationship	288	0.18
88	comput	287	0.18
89	class	286	0.18
90	game	285	0.18
91	context	284	0.17
92	propos	284	0.17
93	identifi	283	0.17
94	internet	283	0.17
95	influenc	280	0.17
96	outcom	279	0.17
97	percept	276	0.17
98	compar	274	0.17
99	network	274	0.17
100	task	269	0.17

Top 100 n-grams for online learning articles 1997-2016

	N_Gram	Total	per Article
1	learn	11802	2.45
2	student	10452	2.17
3	onlin	6827	1.42
4	educ	5184	1.08
5	cours	4325	0.90
6	base	3381	0.70
7	develop	3184	0.66
8	design	2853	0.59
9	technolog	2775	0.58
10	web	2420	0.50
11	learner	2382	0.49
12	result	2352	0.49
13	teacher	2343	0.49
14	environ	2313	0.48
15	discuss	2216	0.46
16	effect	2161	0.45
17	system	2125	0.44
18	distanc	2106	0.44
19	commun	2083	0.43
20	interact	2062	0.43
21	particip	1991	0.41
22	support	1985	0.41
23	teach	1952	0.40
24	model	1778	0.37
25	univers	1768	0.37
26	inform	1745	0.36
27	experi	1744	0.36
08	instruct	1711	0.35
29	knowledg	1557	0.32
30	activ	1541	0.32
31	collabor	1530	0.32
32	social	1486	0.31
33	process	1472	0.31
34	assess	1409	0.29
35	internet	1402	0.29
36	school	1362	0.28
37	practic	1360	0.28
38	data	1359	0.28
39	evalu	1337	0.28
40	tool	1328	0.28
41	analysi	1304	0.27
42	comput	1302	0.27
43	approach	1254	0.26
44	program	1250	0.26
45	level	1239	0.26
46	virtual	1230	0.26
47	provid	1216	0.25
48	perform	1164	0.24
49	self	1157	0.24
50	find	1155	0.24
51	learning_env iron	1095	0.23
52	content	1085	0.23
53	factor	1069	0.22
54	includ	1056	0.22
55	web_bas	1046	0.22
56	online_learn	1022	0.21
57	suggest	1002	0.21
58	relat	988	0.20
59	time	983	0.20
60	distance_ed uc	982	0.20
61	instructor	971	0.20
62	strategi	967	0.20
63	indic	946	0.20
64	studi	916	0.19
65	test	895	0.19
66	institut	889	0.18
67	skill	887	0.18
68	resourc	875	0.18
69	qualiti	847	0.18
70	project	845	0.18
71	method	843	0.17
72	academ	841	0.17
73	improv	837	0.17
74	engag	835	0.17
75	signific	833	0.17
76	differ	829	0.17
77	implement	803	0.17
78	online_cours	801	0.17
79	context	796	0.17
80	question	791	0.16
81	class	783	0.16
82	understand	782	0.16

100 n-grams for online learning articles 1997-2016

83	identifi	778	0.16
84	role	775	0.16
85	classroom	774	0.16
86	motiv	774	0.16
87	survei	774	0.16
88	increas	768	0.16
89	issu	768	0.16
90	train	768	0.16
91	facilit	752	0.16
92	structur	748	0.16
93	offer	743	0.15
94	found	740	0.15
95	cognit	739	0.15
96	peer	733	0.15
97	integr	725	0.15
98	manag	722	0.15
99	help	713	0.15
100	network	713	0.15

APPENDIX E

YEARLY ARTICLE COUNT FOR CLUSTERS

Table E.1

Yearly Cluster Data (C1-C6)

Year	C1	C2	C3	C4	C5	C6
1997	17	15	7	15	3	3
1998	14	16	10	26	2	1
1999	16	18	14	30	3	7
2000	37	25	5	45	18	7
2001	31	25	18	51	5	5
2002	27	18	8	43	8	16
2003	22	10	9	50	15	13
2004	24	16	9	57	10	14
2005	14	14	11	48	14	13
2006	10	13	3	46	11	11
2007	13	13	17	58	10	17
2008	13	18	13	64	11	22
2009	19	13	8	48	16	28
2010	14	15	10	71	11	18
2011	15	19	9	49	9	23
2012	14	20	8	52	24	20
2013	9	12	13	56	9	13
2014	7	10	9	58	21	8
2015	6	16	12	56	16	12
2016	5	8	4	53	9	17
TOTAL	327	314	197	976	225	268
Mean	2004.606	2005.806	2006.315	2007.536	2008.089	2008.201
Median	2003	2005.5	2007	2008	2009	2009
Mode	2000	2000	2001	2010	2012	2009
SD	5.100857	5.666772	5.605431	5.323075	5.20569	4.713166

Table E.2

Yearly Cluster Data (C7-C12)

Year	C7	C8	C9	C10	C11	C12
1997	2	1	2	5	2	1
1998	2	0	7	1	0	3
1999	2	4	2	4	1	2
2000	11	6	9	5	0	1
2001	8	11	9	10	1	0
2002	5	4	7	7	0	1
2003	10	9	3	9	1	1
2004	15	17	5	5	2	0
2005	12	15	6	12	2	4
2006	18	25	5	9	2	1
2007	13	20	7	12	1	1
2008	17	15	20	18	1	1
2009	13	24	20	29	3	3
2010	22	11	24	13	4	5
2011	18	19	17	17	3	1
2012	12	23	18	24	3	4
2013	19	24	12	11	6	5
2014	21	20	11	19	6	7
2015	7	15	13	18	2	7
2016	14	10	14	18	1	2
TOTAL	241	273	211	246	41	50
Mean	2008.465	2008.718	2008.773	2009.045	2009.415	2009.42
Median	2009	2009	2010	2009	2011	2011.5
Mode	2010	2006	2010	2009	2013	2014
SD	4.75918	4.466082	5.009078	4.90502	4.989868	5.700304

Table E.3

Yearly Cluster Data (C13-C18)

Year	C13	C14	C15	C16	C17	C18
1997	7	2	0	0	0	0
1998	6	0	0	0	0	0
1999	9	1	1	1	0	0
2000	20	2	1	1	1	0
2001	20	2	4	0	0	0
2002	21	3	3	2	0	0
2003	23	5	5	0	0	0
2004	29	5	5	3	0	0
2005	33	5	4	5	0	0
2006	37	2	5	6	3	0
2007	52	6	8	9	4	0
2008	52	8	7	5	8	0
2009	80	14	9	4	4	0
2010	68	18	15	10	7	0
2011	59	11	15	10	5	0
2012	84	11	15	19	6	2
2013	81	17	15	6	7	5
2014	70	14	13	5	8	17
2015	88	16	9	11	11	23
2016	61	14	18	8	7	21
TOTAL	900	156	152	105	71	68
Mean	2009.731	2010.378	2010.408	2010.419	2011.592	2014.824
Median	2010	2011	2011	2011	2012	2015
Mode	2015	2010	2016	2012	2015	2015
SD	4.599527	4.405778	4.241634	3.857513	3.353367	1.050106

REFERENCES

AJDE. (2017). Aims and scope. *American Journal of Distance Education*. Retrieved from http://www.tandfonline.com/action/journalInformation?show=aimsScope&journalCode=hajd20

Allen, I. E., & Seaman, J. (2007). *Online nation: Five years of growth in online learning.* Newburyport, MA: Sloan Consortium.

Allen, I. E., & Seaman, J. (2013). *Changing course: Ten years of tracking online education in the United States.* Newburyport, MA: Sloan Consortium.

Anderson, T. (2009). Teaching in an online learning context. In T. Anderson (Ed.), *The theory and practice of online learning* (2nd ed., pp. 343-366). Edmonton, Canada: AU Press, Athabasca University.

Antai, R., Fox, C., & Kruschwitz, U. (2011). The use of latent semantic indexing to cluster documents into their subject areas. *Proceedings of the Fifth Language Technology Conference* (pp. 161-166). New York: Springer.

Archambault, L., & Crippen, K. (2009). Examining TPACK among K-12 online distance educators in the United States. *Contemporary Issues in Technology and Teacher Education, 9*(1), 71-88. https://doi.org/10.13140/2.1.2225.6000

Beaudin, B. P. (1999). Keeping online asynchronous discussions on topic. *Journal of Asynchronous Learning Networks, 3*(2), 41-53.

Benson, A. D. (2002). Using online learning to meet workforce demand: A case study of stakeholder influence. *Quarterly Review of Online Education, 3*(4), 443-452.

Bernard, R. M., Abrami, P. C., Borokhovski, E., Wade, C. A., Tamim, R. M., Surkes, M. A., & Bethel, E. C. (2009). A meta-analysis of three types of interaction treatments in distance education. *Review of Educational Research, 79*(3), 1243-1289. https://doi.org/10.3102/0034654309333844

Bernard, R. M., Abrami, P. C., Lou, Y., Borokhovski, E., Wade, A., Wozney, L., . . . Huang, B. (2004). How does distance education compare with classroom instruction? A meta-analysis of the empirical literature. *Review of Educational Research, 74*(3), 379-439. https://doi.org/10.3102/00346543074003379

Biostat. (2017). Why perform a meta-analysis? Retrieved from https://www.meta-analysis.com/pages/why_do.php

Bledsoe, T. S., & Simmerok, B. D. (2013). A multimedia-rich platform to enhance student engagement and learning in an online environment. *Journal of Asynchronous Learning Networks, 17*(4), 57-66. http://dx.doi.org/10.24059/olj.v17i4.398

Bolkan, J. (2014, June 11). *Report: 83 percent of high schools offer online courses.* Retrieved from https://thejournal.com/Articles/2014/06/11/Report-83-Percent-of-High-Schools-Offer-Online-Courses.aspx?p=1

Bontis, N., & Serenko, A. (2009). A follow-up ranking of academic journals. *Journal of Knowledge Management, 13*(1), 16-26. https://dx.doi.org/10.1108/13673270910931134

Borup, J., West, R. E., & Graham, C. R. (2013). The influence of asynchronous video communication on learner social presence: A narrative analysis of four cases. *Distance Education, 34*(1), 48-63. http://dx.doi.org/10.1080/01587919.2013.770427

Bossu, C., Bull, D., & Brown, M. (2012). Opening up down under: The role of open educational resources in promoting social inclusion in Australia. *Distance Education*, *33*(2), 151-164. https://doi.org/10.1080/01587919.2012.692050

Bowerman, C., Mansfield, C., & Sewell, K. (1997). Using JavaScript to simulate formative assessment questioning in web-based open learning materials. *European Journal of Open, Distance and E-Learning, 1*. Retrieved from http://www.eurodl.org/?p=archives&year=1997&article=1

Bozkurt, A., Akgün-Özbek, E., & Zawacki-Richter, O. (2017). Trends and patterns in massive open online courses: Review and content analysis of research on MOOCs (2008-2015). *The International Review of Research in Open and Distributed Learning, 18*(5). http://dx.doi.org/10.19173/irrodl.v18i5.3080

Bozkurt, A., Akgun-Ozbek, E., Yilmazel, S., Erdogdu, E., Ucar, H., Guler, E., . . . Aydin, C. H. (2015). Trends in distance education research: A content analysis of journals 2009-2013. *The International Review of Research in Open and Distributed Learning, 16*(1). https://doi.org/10.19173/irrodl.v16i1.1953

Carliner, S. (2004). *An overview of online learning* (2nd ed.). Armherst, MA: Human Resource Development Press.

Cavanaugh, C., Gillan, K. J., Kromrey, J., Hess, M., & Blomeyer, R. (2004). *The effects of distance education on K-12 student outcomes: A meta-analysis.* Naperville, IL: Learning Point Associates/ North Central Regional Educational Laboratory (NCREL).

centroid. (2017). In *Merriam-Webster online dictionary*. Retrieved from https://www.merriam-webster.com/dictionary/centroid

Chakraborty, G., Pagolu, M., & Garla, S. (2014). *Text mining and analysis: Practical methods, examples, and case studies using SAS.* Cary, NC: SAS Institute.

Chen, N.-S., Wei, C.-W., & Chen, H.-J. (2008). Mining e-learning domain concept map from academic articles. *Computers & Education*, *50*(3), 1009-1021. https://doi.org/10.1016/j.compedu.2006.10.001

Chen, N.-S., Wei, C.-W., Wu, K.-T., & Uden, L. (2009). Effects of high level prompts and peer assessment of online learners' reflection levels. *Computers & Education, 52*(2), 283-291. http://dx.doi.org/10.1016/j.compedu.2008.08.007

Cheng, Y., & Ye, J. (2010). Exploring the social competence of students with autism spectrum conditions in a collaborative virtual learning environment—The pilot study. *Computers & Education, 54*(4), 1068-1077. http://dx.doi.org/10.1016/j.compedu.2009.10.011

Choi, H. J., & Johnson, S. D. (2005). The effect of context-based video instruction on learning and motivation in online courses. *American Journal of Distance Education, 19*(4), 215-227. http://dx.doi.org/10.1207/s15389286ajde1904_3

Clark, R. C., & Mayer, R. E. (2016). *e-Learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning.* (4th ed.). Hoboken, NJ: John Wiley & Sons.

Clustering: An introduction . (n. d.). Retrieved from https://home.deib.polimi.it/matteucc/Clustering/tutorial_html/

Coban, M., Karakus, T., Karaman, A., Gunay, F., & Goktas, Y. (2015). Technical problems experienced in the transformation of virtual worlds into an education environment and coping strategies. *Journal of Educational Technology & Society*, *18*(1), 37-49.

Conrad, D. (2002). Deep in the hearts of learners: Insights into the nature of online community. *International Journal of E-Learning and Distance Education*, *17*(1), 1-19. Retrieved from http://www.ijede.ca/index.php/jde/article/view/133

Cook, R. G., Ley, K., Crawford, C., & Warner, A. (2008). Motivators and inhibitors for university faculty in distance and e-learning. *British Journal of Educational Technology*, *40*(1), 149-163. http://dx.doi.org/10.1111/j.1467-8535.2008.00845.x

Coomey, M., & Stephenson, J. (2001). Online learning: It is all about dialogue, involvement, support and control—according to the research. In J. Stephenson (Ed.), *Teaching and learning online: Pedagogies for new technologies* (pp. 37-52).

Coronado, R. A., Riddle, D. L., Wurtzel, W. A., & George, S. Z. (2011). Bibliometric analysis of articles published from 1980 to 2009 in Physical Therapy, Journal of the American Physical Therapy Association. *Physical Therapy, 91*(5), 642-655. http://dx.doi.org/10.2522/ptj.20100267

Costa, D. S. J. , Mullan, B. A., Kothe, E. J., & Butow, P. (2010). A web-based formative assessment tool for Masters students: A pilot study. *Computers & Education, 54*(4), 1248-1253. http://dx.doi.org/10.1016/j.compedu.2009.11.011

Croft, S. J., Roberts, M. A., & Stenhouse, V. L. (2016). The perfect storm of education reform: High-stakes testing and teacher evaluation. *Social Justice, 42*(1), 70-92.

Cui, W., Wu, Y., Liu, S., Wei, F., Zhou, M. X., & Qu, H. (2010). Context-preserving, dynamic word cloud visualization. *IEEE Computer Graphics and Applications, 30*(6), 42-53. http://dx.doi.org/10.1109/MCG.2010.102

de Langen, F. H. T., & Bitter-Rijkema, M. E. (2012). Positioning the OER business model for open education. *European Journal of Open, Distance and E-Learning, 15*(1). Retrieved from http://www.eurodl.org/?p=archives&year=2012&half&article=483

Delen, D., & Crossland, M. D. (2008). Seeding the survey and analysis research literature with text mining. *Expert Systems with Applications*, *34*(3), 1707-1720. https://doi.org/10.1016/j.eswa.2007.01.035

Denning, T., Fisher, F., & Higgins, C. (2011). From cradle to brave new world: The first 20 years of developing a research field in new technologies and teacher education, as reflected in the pages of JITTE/TPE. *Technology, Pedagogy, and Education, 20*(3), 263-288. https://doi.org/10.1080/1475939X.2011.610929

Determining the optimal number of clusters. (n.d.). Retrieved from http://www.sthda.com/english/wiki/print.php?id=239

Dickey, M. D. (2005). Three-dimensional virtual worlds and distance learning: two case studies of Active Worlds as a medium for distance education. *British Journal of Educational Technology*, *36*(3), 439-451. http://dx.doi.org/10.1111/j.1467-8535.2005.00477.x

Duncan, H. E., & Barnett, J. (2009). Learning to teach online: What works for pre-service teachers. *Journal of Educational Computing Research*, *40*(3), 357-376. https://doi.org/10.2190/EC.40.3.f

Durmuşoğlu, A. (2016). A pre-assessment of past research on the topic of environmental-friendly electronics. *Journal of Cleaner Production*, *129*, 305-314. https://doi.org/10.1016/j.jclepro.2016.04.068

Educational Technology Magazine. (n d.). Retrieved from http://aect.site-ym.com/page/ETM

Egarievwe, S. U., Ajiboye, A. O., Biswas, G., Okobiah, O. K., Fowler, L. K. A., Thorne, S. K., & Collins, W. E. (2000). Internet application of LabVIEW in computer based learning. *European Journal of Open, Distance, and E-Learning, 3*(2). Retrieved from http://www.eurodl.org/index.php?p=archives&sp=full&article=91

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing, 62*(1), 107-115. https://doi.org/10.1111/j.1365-2648.2007.04569.x

Elsevier. (2016). Scopus. Retrieved from https://www.elsevier.com/solutions/scopus

European Distance and E-Learning Network. (n.d.). Aims and policies. Retrieved from http://www.eurodl.org/?p=about&sp=aims

Evangelopoulos, N., Zhang, X., & Prybutok, V. R. (2012). Latent semantic analysis: Five methodological recommendations. *European Journal of Information Systems, 21*(1), 70-86. https://doi.org/10.1057/ejis.2010.61

Fairlie, R. W., & Bulman, G. (2016). *Technology and education: Computers, software, and the Internet* (NBER Working Paper No. w22237). Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w22237

Feinberg, J. (2014). Wordle [Computer software]. Retrieved from http://www.wordle.net/create

Feldman, R., & Dagan, I. (1995). Knowledge discovery in textual databases (KDT). In U. M. Fayyad, & R. Uthurusamy (Eds.), *Proceedings of the First International Conference on Knowledge Discovery and Data Mining (KDD-95)* (pp. 112-117). Palo Alto, CA: American Association for Artificial Intelligence.

Franklin University. (2016). The growth of online education over the last decade [Web log post]. Retrieved from http://www.franklin.edu/blog/the-growth-of-online-education-over-the-last-decade

Fredericksen, E., Pickett, A., Shea, P., Pelz, W., & Swan, K. (2000). Factors influencing faculty satisfaction with asynchronous teaching and learning in the SUNY learning network. *Journal of Asynchronous Learning Networks, 4*(3), 245-278.

Freidman, J. (2016, February 9). Study: Enrollment in online learning up, except at for-profits. *U.S. News and World Report*. Retrieved from http://www.usnews.com/education/online-education/articles/2016-02-09/study-enrollment-in-online-learning-up-except-at-for-profits

Ganesan, K. (2014, November 23). What are n-grams? Retrieved from http://text-analytics101.rxnlp.com/2014/11/what-are-n-grams.html

Garrison, D. R. (2011). *E-Learning in the 21st century: A framework for research and practice* (2nd ed.). New York: Routledge.

Garrison, S., Fenton, R., & Vaissiere, S. (2001). Determining the impact of training on teacher use for a web-to-database system. *Journal of Educational Technology & Society, 4*(2), 145-173.

Gartner, Inc. (2017). Gartner Hype Cycle. Retrieved from https://www.gartner.com/technology/research/methodologies/hype-cycle.jsp

Geith, C., & Vignare, K. (2008). Access to education with online learning and open educational resources: Can they close the gap? *Journal of Asynchronous Learning Networks*, *12*(1), 105-126.

Goddard, N., & Annaraj, D. (2017). *Conflict sensitivity meta-trends analysis.* Monrovia, CA: World Vision International.

Google Scholar. (2017). Metrics [Database]. Retrieved May 15, 2017, from https://scholar.google.com/citations?view_op=metrics_intro&hl=en

Gunawardena, C. N., & Duporne, P. L. (2000). Predictors of learner satisfaction in an academic computer conference. *Distance Education*, *21*(1), 101-117. http://dx.doi.org/10.1080/0158791000210107

Halverson, L. R., Graham, C. R., Spring, K. J., & Drysdale, J. S. (2012). An analysis of high impact scholarship and publication trends in blended learning. *Distance Education, 33*(3), 381-413. http://dx.doi.org/10.1080/01587919.2012.723166

Han, I., & Shin, W. S. (2016). The use of a mobile learning management system and academic achievement of online students. *Computers & Education*, *102*, 79-89. https://doi.org/10.1016/j.compedu.2016.07.003

Harasim, L. (2000). Shift happens: Online education as a new paradigm in learning. *The Internet and Higher Education, 3*(1), 41-61. https://doi.org/10.1016/S1096-7516(00)00032-4

Harpaz, R., Callahan, A., Tamang, S., Low, Y., Odgers, D., Finlayson, S., . . . Shah, N. H. (2014). Text mining for adverse drug events: The promise, challenges, and state of the art. *Drug Safety, 37*(10), 777-790. https://doi.org/10.1007/s40264-014-0218-z

Hayes, S., Uzuner-Smith, S., & Shea, P. (2015). Expanding learning presence to account for the direction of regulative intent: Self-, co- and shared regulation in online learning. *Online Learning, 19*(3), 15-31. doi: https://doi.org/10.24059/olj.v19i3.666

Hearst, M. A. (1999). Untangling text data mining. *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics* (pp. 3-10). Stroudsburg, PA: Association for Computational Linguistics. https://doi.org/10.3115/1034678.1034679

Hickey, R. (2014, December 05). The history of online education. Retrieved from https://www.petersons.com/college-search/online-education-history.aspx#/sweeps-modal

Hooshyar, D., Ahmad, R. B., Yousefi, M., Fathi, M., Abdollahi, A., Horng, S.-J., & Lim, H. (2016). A solution-based intelligent tutoring system integrated with an online game-based formative assessment: Development and evaluation. *Educational Technology Research and Development*, *64*(4), 787-808. https://doi.org/10.1007/s11423-016-9433-x

Huang, C.-C., Yeh, T.-K., Li, T.-Y., & Chang, C.-Y. (2010). The idea storming cube: Evaluating the effects of using game and computer agent to support divergent thinking. *Journal of Educational Technology & Society, 13*(4), 180-191.

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

Hung, J.-L., & Zhang, K. (2012). Examining mobile learning trends 2003-2008: A categorical meta-trend analysis using text mining techniques. *Journal of Computing in Higher Education, 24*(1), 1-17. https://doi.org/10.1007/s12528-011-9044-9

Hwang, G.-J., & Tsai, C.-C. (2011). Research trends in mobile and ubiquitous learning: A review of publications in selected journals from 2001 to 2010. *British Journal of Educational Technology, 42*(4), E65-E70. https://doi.org/10.1111/j.1467-8535.2011.01183.x

Hwang, G.-J., & Wu, P.-H. (2012). Advancements and trends in digital game-based learning research: A review of publications in selected journals from 2001 to 2010. *British Journal of Educational Technology, 43*(1), E6-E10. https://doi.org/10.1111/j.1467-8535.2011.01242.x

Immerwahr, J. (2004). *Public attitudes on higher education: A trend analysis, 1993 to 2003* (National Center Report #04-2). San Jose, CA: The National Center for Pubic Policy and Higher Education.

International Association for K-12 Online Learning (iNACOL). (2011). *The online learning definitions project.* Vienna, VA: International Association for K-12 Online Learning.

Jaggars, S. S., & Bailey, T. (2010). *Effectiveness of fully online courses for college students: Response to a department of education meta-analysis.* New York, NY: Community College Research Center, Columbia University.

Jain, A. K. (2010). Data clustering: 50 years beyond k-means. *Pattern Recognition Letters, 31*(8), 651-666. https://doi.org/10.1016/j.patrec.2009.09.011

Jo, I.-H., Kim, D., & Yoon, M. (2015). Constructing proxy variables to measure adult learners' time management strategies in LMS. *Journal of Educational Technology & Society, 18*(3), 214-225.

Johnstone, S. M. (2005). Open educational resources serve the world. *Educause Quarterly, 28*(3), 15-18.

Joksimović, S., Kovanović, V., Skrypnyk, O., Gašević, D., Dawson, S., & Siemens, G. (2015). The history and state of online learning. In G. Siemens, D. Gašević, & S. Dawson, *Preparing for the Digital University : A Review of the History and Current State of Distance, Blended, and Online Learning* (pp. 96-131). Alberta, Canada, Alberta, Canada: Athabasca University.

Kabakoff, R. (2017). Cluster analysis. Retrieved from http://www.statmethods.net/advstats/cluster.html

Keegan, D. (1995). *Distance education technology for the new millennium.* Hagen, Germany: Fern University, Institute for Research into Distance Education.

Keshava, Gireesh, A. G., & Gowda, M. P. (2008). ACM transaction on information systems (1989-2006): A bibliometric study. *Information Studies*, *14*(4), 223-234.

Kilis, S., Gülbahar, Y., & Rapp, C. (2016). Exploration of teaching preferences of instructors' use of social media. *European Journal of Open, Distance and E-Learning, 19*(1), 1-18. https://doi.org/10.1515/eurodl-2016-0001

Kim, C. (2012). The role of affective and motivational factors in designing personalized learning environments. *Educational Technology Research and Development, 60*(4), 563-584. https://doi.org/10.1007/s11423-012-9253-6

Kim, S. H., Holmes, K., & Mims, C. (2005). Mobile wireless technology use and implementation: Opening a dialogue on the new technologies in education. *TechTrends, 49*(3), 54-63. https://doi.org/10.1007/BF02763647

Kinshuk, Chen, N.-S., & Sampson, D. G. (2016, December 9). Dear ET&S authors, reviewers, and readers [Letter from editors]. *Journal of Educational Technology & Society*. Retrieved from http://ifets.info/Announcements/1481271385.pdf

Kireyev, K. (2008). Using latent semantic analysis for extractive summarization. *Proceedings of the Text Analysis Conference.* Gathersburg, MD: National Institute of Standards and Technology.

Kovanović, ,. V., Joksimović, S., Skrypnyk, O., Gašević, D., Dawson, S., & Siemens, G. (2015). The history and state of distance education. In G. Siemens, D. Gašević, & S. Dawson, *Preparing for the digital university: A review of the history and current state of distance, blended, and online learning* (pp. 9-54). Alberta, Canada: Athabasca University.

Kovanović, V., Joksimović, S., Gašević, D., Siemens, G., & Hatala, M. (2015). What public media reveals about MOOCs: A systematic analysis of news reports. *British Journal of Educational Technology, 46*(3), 510-527. https://doi.org/10.1111/bjet.12277

Kožuh, I., Jeremić, Z., Sarjaš, A., Bele, J. L., Devedžić, V., & Debevc, M. (2015). Social presence and interaction in learning environments: The effect on student success. *Journal of Educational Technology & Society, 18*(1), 223-326.

Lazard, A. J., Scheinfeld, E., Bernhardt, J. M., Wilcox, G. B., & Suran, M. (2015). Detecting themes of public concern: A text mining analysis of the Centers for Disease Control and Prevention's ebola live Twitter chat. *American Journal of Infection Control, 43*(10), 1109-1111. https://doi.org/10.1016/j.ajic.2015.05.025

Lee, Y., Driscoll, M. P., & Nelson, D. (2004). The past, present, and future of research in distance education: Results of a content analysis. *American Journal of Distance Education*, *18*(4), 225-241. https://doi.org/10.1207/s15389286ajde1804_4

Lextek International. (2002). Stop Word List 1. Retrieved from http://www.lextek.com/manuals/onix/stopwords1.html

Lin, L. (2008). An online learning model to facilitate learners' rights to education. *Journal of Asynchronous Learning Networks, 12*(1), 127-143.

Lo, H.-Y., Liu, G.-Z., & Wang, T.-I. (2014). Learning how to write effectively for academic journals: A case study investigating the design and development of a genre-based writing tutorial system. *Computers & Education, 78*, 250-267. https://doi.org/10.1016/j.compedu.2014.06.007

Manches, A., Duncan, P., Plowman, L., & Sabeti, S. (2015). Three questions about the Internet of things and children. *TechTrends*, *59*(1), 76-83. https://doi.org/10.1007/s11528-014-0824-8

Martens, B., Poeppelbuss, J., & Teuteberg, F. (2011). Understanding the cloud computing ecosystem: Results from a quantitative content analysis. *Wirtschaftsinformatik* (pp. 466-476). Wiesbaden, Germany: Springer Gabler.

Martin, F., Ahlgrim-Delzell, L., & Budhrani, K. (2017). Systematic review of two decades (1995 to 2014) of research on synchronous online learning. *The American Journal of Distance Education, 31*(1), 3-19. https://doi.org/10.1080/08923647.2017.1264807

Matzat, U. (2013). Do blended virtual learning communities enhance teachers' professional development more than purely virtual ones? A large scale empirical comparison. *Computers & Education, 60*(1), 40-51. https://doi.org/10.1016/j.compedu.2012.08.006

McAuley, A., Stewart, B., Siemens, G., & Cormier, D. (2010). *The MOOC model for digital practice.* SSHRS Knowledge Synthesis Grant on the Digital Economy. Retrieved from http://www.edukwest.com/wp-content/uploads/2011/07/MOOC_Final.pdf

Means, B., Toyama, Y., Murphy, R., & Baki, M. (2013). The effectiveness of online and blended learning: A meta-analysis of the empirical literature. *Teachers College Record, 115*(3), 1-47.

Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2009). *Evaluation of evidenced based practices in online learning: A meta-analysis and review of online learning studies.* Washington, DC: U.S. Department of Education.

Merillat, L., & Scheibmeir, M. (2016). Developing a quality improvement process to optimize faculty success. *Online Learning, 20*(3), 159-172. http://dx.doi.org/10.24059/olj.v20i3.977

Miner, G., Delen, D. E., Fast, T., Hill, R., & Nisbet, R. (2012). The seven practice areas of text analytics. In *Practical Text Mining and Statistical Analysis for Non-Structure Text Data Applications* (pp. 29-41). Cambridge, MA: Academic Press.

Mogil, J. S., Simmonds, K., & Simmonds, M. J. (2009). Pain research from 1975 to 2007: A categorical and bibliometric meta-trend analysis of every research paper published in the journal, Pain. *Pain, 142*(1), 48-58. https://doi.org/10.1016/j.pain.2008.11.012

Molinari, D. L. (2004). The role of social comments in problem-solving groups in an online class. *American Journal of Distance Education, 18*(2), 89-101. https://doi.org/10.1207/s15389286ajde1802_3

Moore, J. L., Dickson-Deane, C., & Galyen, K. (2011). e-Learning, online learning, and distance learning environments: Are they the same? *Internet and Higher Education, 14*(2), 129-135. https://doi.org/10.1016/j.iheduc.2010.10.001

Moore, M. G. (1993). Theory or transactional distance. In D. Keegan (Ed.), *Theoretical Principles of Distance Education* (pp. 22-38). New York, NY: Routledge.

Naidu, S. (2014). Looking back, looking forward: The invention and reinvention of distance education. *Distance Education, 35*(3), 263-270. http://dx.doi.org/10.1080/01587919.2014.961671

National Institutes of Health. (2017, April 6). NIH library writing center: impact factor & h-index. Retrieved from http://nihlibrary.campusguides.com/c.php?g=38330&p=244518

Natividad, G. O. (2016). *An analysis of educational technology publications: Who, what and where in the last 20 years* (Doctoral dissertation). Retrieved from https://digital.library.unt.edu/ark:/67531/metadc849761/m1/1/

Nevgi, A., Virtanen, P., & Niemi, H. (2006). Supporting students to develop collaborative learning skills in technology-based environments. *British Journal of Educational Technology, 37*(6), 937-947. https://doi.org/10.1111/j.1467-8535.2006.00671.x

n-gram. (n. d.). In *Definitions.* Retrieved from http://www.definitions.net/definition/N-GRAM

Ohene-Djan, J., Zimmer, R., Gorle, M., & Naqvi, S. (2003). A personalisable electronic book for video-based sign language education. *Journal of Educational Technology & Society, 6*(4), 86-99.

Online Learning Consortium. (2017). About . Retrieved from https://onlinelearningconsortium.org/read/journals/

Online Learning Consortium. (2017). Our history. Retrieved from https://onlinelearningconsortium.org/about/history/

Open and Distance Learning Association of Australia, Inc. (2017). Aims and scope. *Distance Education*. Retrieved from http://www.tandfonline.com/action/journalInformation?show=aimsScope&journalCode=cdie20

Open University. (2017). Aims and scope. *Open Learning: The Journal of Open, Distance and e-Learning*. Retrieved from http://www.tandfonline.com/action/journalInformation?show=aimsScope&journalCode=copl20

Paulsen, M. F. (2002). *Online education systems: Discussion and definition of terms.* Oslo, Norway: NKI Distance Education.

Picciano, A. G. (2015). How meta-analysis can inform online and blended learning research. In C. D. Dziuban, A. G. Picciano, C. R. Graham, & P. D. Moskal, *Conducting Research in Online and Blended Learning Environments: New Pedagogical Frontiers* (pp. 43-54). London, UK: Routledge.

Prabha, S., Duraiswamy, K., & Sharmila, M. (2016). Analysis of different clustering techniques in data and text mining. *International Journal of Computer Science Engineering (IJCSE), 3*(2), 107-116.

Pushplata, R. C. (2012). An analytical assessment on document clustering. *International Journal of Computer Network and Information Security, 4*(5), 63-71. https://doi.org/10.5815/ijcnis.2012.05.08

Quain, M., & Clark, K. R. (2016). Game-Based learning and 21st century skills: A review of recent reserach. *Computers and Human Behavior, 63*, 50-58. https://doi.org/10.1016/j.chb.2016.05.023

R Foundation. (2017). R project [Software]. Retrieved from https://www.r-project.org/

Ragoonaden, K., & Bordeleau, P. (2000). Collaborative learning via the Internet. *Journal of Educational Technology & Society, 3*(3), 361-372.

Ramos, J. (2003). Using tf-idf to determine word relevance in document queries. *Proceedings of the First Instructional Conference on Machine Learning* (pp. 133-142). Piscataway, NJ: Rutgers University.

RapidMiner. (2017). Rapid miner studio 7.6 [Computer software]. Retrieved from https://rapidminer.com/

Reeve, F., Gallacher, J., & Mayes, T. (1998). Can new technology remove barriers to work-based learning? *Open Learning: The Journal of Open, Distance and e-Learning, 13*(3), 18-26.

Reeves, T. C. (2000). Alternative Assessment Approaches for Online Learning Environments in Higher Education. *Journal of Educational Computing Research, 23*(1), 101-111. https://doi.org/10.2190/GYMQ-78FA-WMTX-J06C

Reilly, E. D., Williams, K. M., Stafford, R. E., Corliss, S. B., Walkow, J. C., & Kidwell, D. K. (2016). Global times call for global measures: Investigating automated essay scoring in linguistically-diverse MOOCs. *Online Learning, 20*(2), 217-229. http://dx.doi.org/10.24059/olj.v20i2.638

Rodriguez, C. O. (2012). MOOCs and the AI-Stanford like courses: Two successful and distinct course formats for massive open online courses. *European Journal of Open, Distance and E-Learning, 15*(2). Retrieved from http://www.eurodl.org/index.php?p=archives&year=2013&halfyear=2&article=516

Rogers, A. (2003). *What is the difference?: A new critique of adult learning and teaching.* Oxford: National Institute of Adult Continuing Education.

Roulston, K., & Shelton, A. S. (2015). Reconceptualizing bias in teaching qualitative research methods. *Qualitative Inquiry, 21*(4), 332-342. https://doi.org/10.1177/1077800414563803

Scapin, R. H., & Marega, E. J. (2000). A new tool for managing students' questions in web-bsed distance education courses. *American Journal of Distance Education, 14*(3), 49-57. http://dx.doi.org/10.1080/08923640009527064

Schlosser, L. A., & Simonson, M. (2003). Defining distance education [pdf file]. Retrieved September 18, 2017, from http://www.aect.org/Intranet/Publications/Disted/Disted.asp

SCImago. (2007). SJR-SCImago journal & country rank [Database portal]. Retrieved May 15, 2017, from http://www.scimagojr.com

Scott, K. M. (2013). Does a university teacher need to change e-learning beliefs and practices when using a social networking site? A longitudinal case study. *British Journal of Educational Technology, 44*(4), 571-580. https://doi.org/10.1111/bjet.12072

Siemens, G. (2013). Massive open online courses: Innovation in education? In R. McGreal, W. Kinuthis, & S. Marshall (Eds.), *Open educational resources: Innovation, research and practice* (pp. 5-15).

Simonson, M., Schlosser, C., & Orellana, A. (2011). Distance education research: A review of the literature. *Journal of Computing in Higher Education, 23*(2), 124-142. https://doi.org/10.1007/s12528-011-9045-8

Simpson, O. (2013). Student retention in distance education: Are we failing our students? *Open Learning: The Journal of Open, Distance, and e-Learning, 28*(2), 105-119. https://doi.org/10.1080/02680513.2013.847363

Skogs, J. (2013). Subject line preferences and other factors contributing to coherence and interaction in student discussion forums. *Computers & Education, 60*(1), 172-183. https://doi.org/10.1016/j.compedu.2012.07.005

Sloane, A. (1997). Learning with the Web: Experience using the World Wide Web in a learning environment. *Computers & Education, 28*(4), 207-212. https://doi.org/10.1016/S0360-1315(97)00017-1

Smith, A. (2015, April 21). The increasingly digital community college. Retrieved September 14, 2017, from https://www.insidehighered.com/news/2015/04/21/survey-shows-participation-online-courses-growing

Söder, O. (2008, May 29). How does k-means clustering work? Retrieved from http://www.fon.hum.uva.nl/praat/manual/k-means_clustering_1__How_does_k-means_clustering_work_.html

Spatariu, A., Hartley, K., Schraw, G., Bendixen, L. D., & Quinn, L. F. (2007). The influence of the discussion leader procedure on the quality of arguments in online discussions. *Journal of Educational Computing Research, 37*(1), 83-103. https://doi.org/10.2190/B580-2R37-P220-65V4

Spector, J. M. (2007). Competencies for online teachers. In J. M. Spector (Ed.), *Finding your online voice: Stories told by experienced online educators* (pp. 1-18). New York, NY: Routledge.

Spector, J. M., Ikenthaler, D., Sampson, D., Yang, L., Mukama, E. W., Dona, K. L., . . . Gibson, D. (2016). Technology enhanced formative assessment for 21st century learning. *Journal of Educational Technology & Society*, *19*(3), 58-71.

Spector, J. M., Johnson, T. E., & Young, P. A. (2014). An editorial on research and development in and with educational technology. *Educational Technology, Research and Development*, *62*(1), 1-12. https://doi.org/10.1007/s11423-014-9331-z

Stegmann, K., Wecker, C., Weinberger, A., & Fischer, F. (2012). Collaborative argumentation and cognitive elaboration in a computer-supported collaborative learning environment. *Instructional Science*, *40*(2), 297-323. https://doi.org/10.1007/s11251-011-9174-5

Stemler, S. (2001). An overview of content analysis. *Practical Assessment Research & Evaluation, 7*(17), 137-146.

Straumsheim, C. (2017, May 2). Volatile but growing online ed market. Retrieved from https://www.insidehighered.com/news/2017/05/02/report-finds-growth-volatility-online-education-market

Takacs, J., Reed, J. W., Wells, J. G., & Dombrowski, L. A. (1999). The effects of online multimedia project development, learning style, and prior computer experiences on teachers' attitudes toward the Internet and hypermedia. *Journal of Research on Computing in Education, 31*(4), 341-355. https://doi.org/10.1080/08886504.1999.10782259

Taylor, J. C. (2001). Fifth generation distance education. *Instructional Science and Technology, 4*(1), 1-14.

Thelwall, M. (2008). Bibliometrics to webometrics. *Journal of Information Science, 34*(4), 605-621. https://doi.org/10.1177/0165551507087238

Tirovolas, A. K., & Levitin, D. J. (2011). Music perception and cognition research from 1983 to 2010: A categorical and bibliometric analysis of empirical articles in Music Preparation. *Music Preparation, 29*(1), 23-36. https://doi.org/10.1525/mp.2011.29.1.23

trend. (2016). In *Merriam-Webster's online dictionary.* Retrieved from https://www.merriam-webster.com/dictionary/trend

trend analysis. (2017). In *Business dictionary online.* Retrieved from http://www.businessdictionary.com/definition/trend-analysis.html

Trespalacious, J., & Rand, J. (2015). Using asynchronous activities to promote sense of community and learning in an online course. *International Journal of Online Pedagogy and Course Design, 5*(4), 1-13. https://doi.org/10.4018/IJOPCD.2015100101

Tsai, C.-W., Shen, P.-D., & Fan, Y.-T. (2013). Research trends in self-regulated learning research in online learning environments: A review of studies published in selected journals from 2003 to 2012. *British Journal of Educational Technology, 44*(5), E107-E110. https://doi.org/10.1111/bjet.12017

Tsai, M.-J. (2009). Online information searching strategy inventory (OISSI): A quick verson and a complete version. *Computers & Education, 53*(2), 473-483. https://doi.org/10.1016/j.compedu.2009.03.006

Veletsianos, G., & Kimmons, R. (2012). Networked participatory scholarship: Emergent techno-cultural pressures toward open and digital scholarship in online networks. *Computers & Education, 58*(2), 766-774. https://doi.org/10.1016/j.compedu.2011.10.001

Volery, T., & Lord, D. (2000). Critical success factors in online education. *The International Journal of Educational Management, 14*(5), 216-223.

Watson, J. B., & Rossett, A. (1999). Guiding the independent learner in Web-based training. *Educational Technology, 39*(3), 27-36.

Watson, J. F., & Kalmon, S. (2005). *Keeping pace with K-12 online learning: Review of state-level policy and practice.* Naperville, IL: Learning Point Associates.

Welsch, E. T., Wanberg, C. R., Brown, K. G., & Simmering, M. J. (2003). E-Learning: Emerging uses, empirical results and future directions. *International Journal of Training and Development, 7*(4), 245-258. https://doi.org/10.1046/j.1360-3736.2003.00184.x

White, G. O., Guldiken, O., Hemphill, T. A., He, W., & Khoobdeh, M. S. (2016). Trends in international strategic management research from 2000 to 2013: Text mining and bibliometric analyses. *Management International Review, 56*(1), 35-65. https://doi.org/10.1007/s11575-015-0260-9

Wiemer-Hastings, P., Wiemer-Hastings, K., & Graesser, A. (2004). Latent semantic analysis. *Proceedings of the 16th International Joint Conference on Artificial Intelligence* (pp. 1-14). San Francisco, CA: Morgan Kaufman.

Wikeley, F., & Muschamp, Y. (2007). Pedagogical implications of working with docoral students at a distance. *Distance Education*, *25*(1), 125-142. https://doi.org/10.1080/0158791042000212495

Willett, P. (2006). The porter stemming algorithm: Then and now. *Program*, *40*(3), 219-223. https://doi.org/10.1108/00330330610681295

Williams, D., Ma, Y., Feist, S., Richard, C. E., & Prejean, L. (2007). The design of an analogical encoding tool for game-based virtual learning environments. *British Journal of Educational Technology, 38*(3), 429-437. https://doi.org/10.1111/j.1467-8535.2007.00707.x

Wong, Y. Y., Zeng, J., & Ho, C. K. (2016). Trends in open and distance learning research: 2005 vs 2015. *Asian Association of Open Universities Journal*, *11*(2), 216-177. https://doi.org/10.1108/AAOUJ-09-2016-0035

Yang, J., & Chen, X. (2002). A semi-structured document model for text mining. *Journal of Computer Science and Technology, 17*(5), 603-610. https://doi.org/10.1007/BF02948828

Yim, S., & Warschauer, M. (2017). Web-based collaborative writing in L2 contexts: Methodological insights from text mining. *Language Learning & Technology, 21*(1), 146-165.

Yu, C. H., Jannasch-Pennell, A., & DiGangi, S. (2011). Compatibility between text mining and qualitative research in the perspectives of grounded theory, content analysis, and reliability. *The Qualitative Report, 16*(3), 730-744.

Zawacki-Richter, O., & Naidu, S. (2016). Mapping research trends from 35 years of publications in distance education. *Distance Education, 37*(3), 245-269. https://doi.org/10.1080/01587919.2016.1185079

Zawacki-Richter, O., Bäcker, E. M., & Vogt, S. (2009). Review of distance education research (2000 to 2008): Analysis of research areas, methods, and authorship patterns. *International Reviw of Research in Open and Distance Learning, 10*(6), 21-50. http://dx.doi.org/10.19173/irrodl.v10i6.741

Zhang, P. (1998). A case study on technology use in distance learning. *Journal of Research on Computing in Education, 30*(4), 398-419. https://doi.org/10.1080/08886504.1998.10782235

Zhu, F., Patumcharoenpol, P., Zhang, C., Yang, Y., Chan, J., Meechai, A., . . . Bairong, S. (2013). Biomedical text mining and its applications in cancer research. *Journal of Biomedical Informatics, 46*, 200-211. https://doi.org/10.1016/j.jbi.2012.10.007