The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/0737-8831.htm

Applied bibliometrics and information visualization for decision-making processes in higher education institutions

Carlos Vílchez-Román
Pontificia Universidad Católica del Perú (PUCP), Lima, Peru and CENTRUM Católica Graduate Business School (CCGBS), Lima, Peru
Sol Sanguinetti
Environment and Technology Institute, Lima, Peru, and
Mariela Mauricio-Salas
Universidad de San Martín de Porres (USMP), Lima, Peru

Received 23 October 2019
Revised 10 February 2020
19 March 2020
Accepted 19 March 2020

Abstract

Purpose – The purpose of this paper is to analyse how using bibliometrics and information visualization can provide a "picture at glance" from which decision-makers can structure processes, thus organizing outputs/outcomes from inception.

Design/methodology/approach – This study carried out a bibliometric-oriented review on studies on higher education students' retention; 1,962 records were downloaded from Scopus and grouped into three five-year intervals: 2002–2006 ($n = 236$), 2007–2011 ($n = 584$) and 2012–2016 ($n = 1{,}142$). Centrality measures and text-mining techniques were used to analyse data.

Findings – Clusters of academic networks were identified by using co-citation analysis. Also, terms with high semantic value were ranked and grouped by using automatic term extraction and co-word analysis.

Practical implications – The bibliometric approach used in this study identifies clusters of authors specialized in student retention, as well as detects the primary terms within this research field.

Originality/value – This paper provides evidence that a bibliometric approach in conjunction with data visualization can be a valuable complement to in-depth literature reviews for the decision-making process.

Keywords Bibliometrics, Information visualization, Decision-making for higher education

Paper type Research paper

Introduction

Student satisfaction with higher education services and its outcomes (e.g. class participation, grades, employability, etc.) is widely studied in academic literature (Gerdes and Mallinckrodt, 1994; Nagda *et al.*, 1998; Thomas, 2002; Tinto, 2006). However, there is a lack of in-depth examination of satisfied students' actions (e.g. student loyalty, student persistence, etc.) and how these affect the outcomes for higher education institutions and students themselves.

Decision-making processes in higher education institutions can result in beneficial academic and organizational outputs; therefore, it is important to examine them to understand and strategically plan to improve higher education centres; consequently, making students more satisfied with the provision of educational services and better professionals.

Within the knowledge discovery process, the identification of patterns, conglomerates or rules of association and classification is one of the final stages of that process, both in the six-

The authors disclose the reception of the following financial support for the research, authorship, and/or publication of this article: This study was supported by CENTRUM Católica Graduate Business School (CCGBS) and the Universidad de San Martín of Porres (USMP).

Library Hi Tech

0737-8831
DOI 10.1108/LHT-10-2019-0209

stage model (Cios and Kurgan, 2005) and in the *n*-stage model (Cios *et al.*, 2007). Information visualization techniques, which are generally based on the map metaphor for representing these structures, are used to identify them in numerical, textual, temporal or spatial data. Unlike traditional graphics, which use the frequency counting of the observed elements, map visualization considers the similarities and differences between the analysed elements. Since maps are generally interactive, users can easily manage the representation of information.

These maps incorporate spatial representation of algorithms, such as cluster analysis and multidimensional scaling, to achieve useable visualizations. In this sense, instead of just exploring numerical or textual data, the visualization of information provides patterns and structures that are easy to identify by decision-makers, researchers and academics in higher education. In this way, the knowledge discovery is completed, its final stage is the use of discovered knowledge, as occurs when decision-makers set aside theoretical models or alternatives with less support and focus for those with greater support or empirical evidence. The latter is observed when the visualization of the information identifies the most influential authors within a field of knowledge and allows for the selection of the most relevant works in specific areas or topics, thereby achieving more efficient use of time within the decision-making process (Chen, 2019).

The visualization of information can be a management support tool, where three main functions have been identified for this visual representation approach: communication medium, knowledge management medium and instrument to support decisions (Al-Kassab *et al.*, 2014).

In the decision-making process, modelling with structural equations has shown that the quality of visualization reduces the complexity of the problem and improves its understanding, which improves the quality of the content (Zabukovec and Jaklič, 2015).

From the perspective of visual rhetorical analysis, which goes beyond the representation of numerical or textual data, design principles have been identified to help to efficiently manage ambiguity and complexity when designing dashboards or business reports (Quattrone, 2017).

In the field of administration, several case studies demonstrated the effectiveness of the information visualization for reducing uncertainty when analysing financial issues (Daradkeh and Abul-Huda, 2017) or sales transactions of small producers (Hoelscher and Mortimer, 2018).

While in the field of decision-making in higher education, an exploratory study conducted with 34 professors and officials of a university showed that the use of visualization criteria facilitated reading and understanding of the presented data (Zentner *et al.*, 2019).

A more recent study analysed the data generated from 2012 to 2018 at the Israel Technological University of Ecuador. The techniques used to represent the information allowed making appropriate decisions to improve the processes in deficient areas (Baldeón *et al.*, 2020).

Decision-makers in higher education institutions can make the most of information visualization, if these tools accurately record all the necessary elements to replicate the studies and their methods. The visualization of the information must include all the stages, from the search, download and extraction of the information that will be processed, to the data analysis and the generation of visual representations.

Therefore, this study outlines a replicable process to achieve information visualization and thus answer the research question: Can visualizing author and thematic clusters provide a better idea on how to make decisions on student retention?

Bibliometrics and text mining approach

Due to the availability and simplification of computer programs, from the 1990s onwards, researchers from different disciplines have relied more on bibliometrics and text-mining-based techniques and approaches for information visualization to improve literature reviews (Delen and Crossland, 2008; Moro *et al.*, 2015; Rodrigues *et al.*, 2014).

Since the research question focusses on thematic and author conglomerates, we used two analytical approaches to identify the intellectual and conceptual structure of student retention studies: co-citation analysis, from bibliometrics and content analysis, from text mining, based on co-occurrence of words and automatic extraction of terms.

Graphics with high visual impact, like those produced with co-citation and co-word analysis, provide significant input for management and decision-making processes (Shang and Wang, 2018; Vílchez-Román *et al.*, 2019; Wetzstein *et al.*, 2019). Nevertheless, for this to work properly, additionally to this tool, it is key to estimate the magnitude of the relationships represented within these graphics. This can be achieved with the measures of centrality and semantic contribution provided this specific type of analytic strategy.

The usefulness of the approach based on text mining tools improves the quality of decision-making, as reported in the academic literature, in areas as diverse as the personalization of online products (Ittoo *et al.*, 2006), the use of medical records for radio-diagnosis (Claster *et al.*, 2008), financial risk management (Lu *et al.*, 2010) and the prioritization of policies to improve governance (Oe *et al.*, 2016).

We selected these strategies for this study because they are complementary and easy to access; by complementary we mean that both approaches let us identify the conceptual and intellectual structure of research in student retention, and by easy to access we mean that our analytical approach can be implemented using free software that have a friendly user interface, based on selection menus, which accelerates the learning curve of text mining tools and information visualization.

Analysis of clusters of authors

Co-citation is an approach that detects clusters of authors. It is based on the number of times two authors, scholar A, the author of a first document, and scholar B, the author of a second document, are cited simultaneously by another author (scholar C) on a third document. This analytical approach follows this chain of reasoning: if two authors, A and B, are cited by a third author, C, it must be because there is a conceptual or theoretical similarity among all of them. For this reason, the three authors appear together when using visualization tools and automatically generating author clusters.

This bibliometric technique gauges the degree of semantic similarity between two or more documents, considering how frequently two authors' publications are co-cited by a third author's publication. The premise behind it is that if there are two documents citing a third one, it is because all of them share a common theoretical or methodological background (Marshakova-Shaikevich, 1973; Small, 1973). Consequently, co-citation analysis helps to detect similarities among a set of academic publications and reveals a latent structure, which was not visible at first glance.

However, this hidden pattern emerges when we complement the co-citation procedure with the visualization tools to identify clusters of authors. At the end of the process, we can see the social structure from which scholars from different disciplines organize themselves, making it possible to identify which ones lead key research agendas (generally, those located in the nuclear zone of the co-citation map) and those who receive little recognition from their academic partners (mostly, located in the peripheral zone of the map).

Content analysis using text-mining techniques

Besides identifying clusters of authors based on co-citation analysis, it is also important to examine the contents of academic publications, using automated approaches, such as text mining tools, because the purpose of the automatic term extraction is to identify the multiword terms with the highest semantic contribution. This is obtained by calculating the C-value score, which is computed by linguistic and statistical procedures, designed for each language. Hence, the analysis generates a ranking of multiword terms that detects the main topics examined in a set of documents. On the other side, clustering and visualization algorithms apply statistical techniques, like cluster analysis or multidimensional scaling, for identifying clusters of authors and keywords to organize contents within a discipline.

This kind of automatically generated content analysis requires structured text to produce an understandable output; therefore, most of the studies that used text mining tools for analysing scientific publications utilized multidisciplinary databases such as Google Scholar, Web of Science (WoS) and Microsoft Academic. Regarding specific text mining tools, some studies have used co-word analysis (Cho, 2014; Surjandari *et al.*, 2015;Vílchez-Román and Quiliano-Terreros, 2017), automatic term extraction algorithms (Rubtsova, 2014; Verbene *et al.*, 2016; Vílchez-Román and Alhuay-Quispe, 2016), while others have worked with clustering and visualization algorithms (Pinto, 2015; Poreau, 2015; Zhu *et al.*, 2013).

Method

Documents source

The three criteria to use Scopus in this study were: (1) this database includes approximately twice as many journals as of WoS, another well-known multidisciplinary database (McCullogh, 2019). As of September 2017, Elsevier, Scopus' publisher, announced that this database reached close to 70m indexed documents, more than 12m author profiles and 70,000 institutional profiles. (2) Scopus also includes indexed sources from non-English-speaking countries, and including these as part of our study was a significant contribution because non-English publications amount between 25 and 30% of academic literature in some disciplines (Aksnes and Sivertsen, 2019; Vera-Baceta *et al.*, 2019). (3) Previous studies have supported the advantages of using Scopus rather than WoS bibliometric database analysis (Mongeon and Paul-Hus, 2016; Moya-Anegón *et al.*, 2007).

Rationale behind the selected computer programs

To examine the conceptual and intellectual structure of student retention research, it was necessary to work with visualization programs that allowed: (1) producing visual representations based on the similarities and differences among the analysed elements, (2) identifying the position of each analysed element (key terms and authors to represent the intellectual structure), based on indicators known as centrality measures. It was necessary to use an application to calculate the contribution or the semantic value of each term in order to measure the importance of the key terms used to describe the context of the research. In mining, one of the indicators to account for such semantic value is known as the C score.

We selected VOSviewer 1.6.5 because it is a free software that uses effective clustering algorithms that generate easily comprehensible data visualization diagrams and identify latent structures within the grouping. As the output is easily understandable, even for non-experts on the data visualization field, it can be used as a tool for a decision-making process for higher education institutions. VOSviewer is available for free download from http://www.vosviewer.com/download

Then we decided to use Pajek 64 4.07 because it is a free tool that quickly calculates of the most known measures of centrality, such as proximity and intermediation. Since it provides such centrality indicators, Pajek is often considered an excellent complement to VOSviewer. Pajek is available for free download from http://mrvar.fdv.uni-lj.si/pajek/

Finally, we chose TerMine because it has the required components of every automatic term-extraction tool (e.g. part-of-speech tagger or algorithm for computing and ranking the obtained *C*-values), and there are many text-mining-based studies that applied this online tool and are indexed in Google Scholar (3,590 mentions in our last search on this academic search engine). TerMine is available from NacTeM website: http://www.nactem.ac.uk/software/termine/

Data collection and information extraction

We defined a 15-year sample period to get a significant overview of the main authors and topic clusters, because the humanities and the social sciences citations take more time to reach their performance peak (Campanario, 2011; Dorta-González and Dorta-González, 2013; Waltman, 2016). We started from the premise that we needed a good volume of studies to conduct this type of analysis; thus, we started this segmentation in 2002 where a critical mass of studies inputted in databases had been reached.

For the analysis, we divided these 15 years into three five-year intervals: 2002–2006, 2007–2011 and 2012–2016, to detect variation in studies for each interval. Our approach was to divide the sample in segments to more easily identify changes from one to the next, taking into account the results of studies that examined the citation impact for each five-year period, and found that the impact factor for five years produced stable and reliable results (Jacsó, 2009, 2010; Liu and Fang, 2020).

For each five-year period, we used the advanced search option available in Scopus with the following search strategy: TITLE-ABS-KEY-AUTH ("student satisfaction" AND ("higher education" OR "tertiary education" OR university OR college) OR (retention OR "retention model" OR "retention program" OR "retention strategies" OR "retention experiences" OR "student retention"))AND (PUBYEAR > [2001 | 2006 | 2011] AND PUBYEAR < [2007 | 2012 | 2017]). After carrying out this process, we downloaded three comma-separated-values (CSV) files for each period: 2002–2006 (n = 236), 2007–2011 (n = 584) and 2012–2016 (n = 1,141). When we analysed sequentially, the samples showed almost a 100% increase for each five-year period. This supports our initial premise of starting when a critical mass was reached.

We extracted four fields for every downloaded record: author, author keywords and index keywords and references. These were selected because they encompass the academic content of publications. The index keyword may vary from authors' as it is assigned by the indexing services of the individual database that hosts each journal; for this reason, we differentiated between author and index keywords.

The authors' data were complete for the three periods; however, there were missing data for the other two fields. In the case of the author's keywords, 134 records (57% of the original data) were identified for the period 2002–2006, 423 records (72% of the original) for 2007–2011 and 966 records (85% of the original) for 2012–2016. In the case of the index keywords, the percentages were 69% (n = 162), 57% (n = 330) and 45% (n = 517) for the periods 2002–2006, 2007–2011 and 2012–2016, respectively. Despite the absence of data for the authors' keywords and index, the selected software considered all the terms available in both fields (which in turn allowed calculating the values of centrality for these terms) when generating the visual representations.

Given that we used the entirety of records obtained from the search strategy detailed earlier (i.e. we did not apply any inclusion or exclusion criteria), this is a census study, as there is complete information from said downloaded records.

Bibliometric analysis and text mining

We used three analytic strategies for the bibliometric analysis and text mining: (1) centrality measures for co-cited authors listed in the references and keywords listed in author and index keywords, (2) automatic term extraction and co-word analysis and (3) visualization of clusters of authors and keywords. The centrality measures allowed us to identify the co-cited authors and the most influential terms, as well as those that connected independent conglomerates.

On the one hand, with the extraction of terms, we obtained a ranking of the semantic value of the keywords that describe the content of each research. Differently, the visualization of clusters of co-cited authors and terms easily organized these elements, avoiding cognitive overload. In that sense, taking into account Miller's findings on the ability to process information through our memory (Braun *et al.*, 1956), we considered that the range between three and five conglomerates was appropriate to easily understand the representation generated with the information display program.

Centrality measures for authors and keywords. For computing centrality measures, we created one diagram for co-cited authors and one for keywords for each period (two diagrams per period x three periods = six total diagrams), then we exported six net files, which enabled data visualization from VOSviewer 1.6.5 and imported them into Pajek64 4.07 for calculating centrality indicators using the features: Network →Create vector → Centrality and selecting the betweenness or closeness options. We obtained the betweenness and closeness score for co-cited authors and keywords in the three five-year periods, in separate diagrams as indicated before.

For this study we used the designation of betweenness as a property of actors (authors) within a social networking structure, where each is represented by a node and the relationships as lines connecting each node. This tells us the degree to which the author functions as a bridge or connector for the remaining authors within the specific social networking structure. In contrast, for closeness we utilized the designation of measurement as the average distance among authors, to determine how adjoining the relationship is with other authors within each specific social networking structure.

Automatic term extraction. We processed both author and index keywords following the same process for both, with the difference that author keywords use co-citation while index keywords use co-word analysis. We applied this process in order to obtain the *C*-value for each multiword term, as this score measures the semantic contribution of a set of multiword terms within a set of documents and provides meaning within the analysed text.

We carried out this process with TerMine (Frantzi *et al.*, 2000). The query interface of the TerMine generates a descending ranking of multiword terms placing on the top the highest *C*-value score. In this way we quickly identified the most relevant topics within a set of documents, since the *C*-value score is obtained by applying statistical and linguistic analysis techniques which significantly reduce the possibility of a bias.

Visualization of authors and keywords clusters. We imported the three five-year period CSV files into VOSviewer 1.6.5 for obtaining the actual mapping of authors and keywords clusters.

For the authors, we used the co-citation option in the "type of analysis" menu and cited authors in the "unit of analysis" menu. We defined a threshold of at least 3, 10 and 15 citations for 2002–2006, 2007–2011 and 2012–2016, respectively, to be included in the clusters of authors.

For the keywords, we used the co-occurrence option in the "type of analysis" menu, all keywords in the "unit of analysis menu" and 3, 10 and 8 as the threshold for the minimum number of occurrences in each of the periods. We worked with the default settings for normalization and clustering algorithms.

The number of citations we determined for each period was necessary to obtain data visualization imaging with 300–400 nodes which allowed us to identify the subjacent structure within nodes. The variation in the number of citations had a direct effect on the visual representation, when the citations were too small, and the nodes were too few for the relationships to be apparent, and when they exceeded the 500 nodes, the subjacent structure was not identifiable. We obtained these numbers by trial and error.

Table 1.
Ranking of authors with higher centrality values (2002–2016)

Rank	Author	Closeness	Betweenness
1	Anderson, R.E.	0.9242	0.0064
2	Black, W.C.	0.9242	0.0067
3	Hair, J.F.	0.9208	0.0067
4	Tatham, R. L	0.9071	0.0062
5	Oliver, R.L.	0.9004	0.0049
6	Tinto, V.	0.8841	0.0055
7	Parasuraman, A	0.8809	0.0045
8	Nunnally, J. C	0.8746	0.0061
9	Zeithmal, V.A.	0.8714	0.0041
10	Babin, B.J.	0.8683	0.0050
11	Fornell, C.	0.8683	0.0043
12	Berry, L.L.	0.8592	0.0034
13	Marsh, H.W.	0.8592	0.0099
14	Davies, J.	0.8443	0.0034
15	Ramsden, P.	0.8443	0.0064

Note(s): Centrality values for terms were computed with the software Pajek64
Source(s): Created by the authors, using results from study

Results

Centrality measures for co-cited authors

Betweenness centrality scores changed from each five-year period to the next. They provided a general view of the authors' influence, for example, Herbert W. Marsh, William Cormack Black or Joseph F. Hair, who acted as gatekeepers and to a lesser extent, as the closest academic references (see Table 1). We use these designations because those authors connect with other areas or topics, but they do not appear among the most frequently cited references. For example, Professor Herbert W. Marsh is one of the most cited educational psychologists in the world, and he appears as a connecting node due to his studies focussed on the students' evaluations of university teaching effectiveness. On his side, Joseph F. Hair appears as another connecting element due to his extensive academic production on multivariate analysis techniques which, although are not directly related to the student satisfaction or the retention strategies and programmes, do provide analytical tools to understand how these two variables interact.

Three small tables for each period are included in the Appendix 1.

Text mining and centrality measures of keywords

The algorithm for automatic term extraction regarding the most important research on student retention showed a medium level of stability. We observed some level of transformation among the terms between five-year periods; however, the top five terms with higher centrality remained unchanged through the three five-year periods (see Table 2).

Three small tables for each period are included in the Appendix 1.

Visualization of clusters of co-cited authors and keywords

Even though we observed changing patterns from one five-year period to the next, at the aggregate level (15 years in total) we found five clusters of authors and six clusters of keywords (see Figures 1 and 2) that let us identify patterns within research studies on student retention. As of authors' clusters, we assigned authors with higher centrality values to each of them (see Table 3).

Table 2.
Multi-word terms with higher *C*-values and centrality values (2002–2016)

Rank	Term	*C*-value	Closeness	Betweenness
1	Student satisfaction	845.604	0.9968	0.0283
2	Medical education	290.000	0.8414	0.0066
3	Medical student	230.679	0.8259	0.0057
4	Higher education	212.081	0.7212	0.0103
5	Programme evaluation	104.222	0.7864	0.0055
6	Educational measurement	95.000	0.7825	0.0045
7	Service quality	93.440		
8	Personal satisfaction	89.000	0.7709	0.0040
9	Education programme	89.000	0.7904	0.0048
10	Student attitude	87.000	0.7579	0.0041
11	Problem-based learning	87.000	0.7279	0.0036
12	Psychological aspect	72.000	0.7330	0.0029
13	Online learning	70.448		
14	Distance education	69.250		0.0016
15	Customer satisfaction	67.561		

Note(s): Centrality values for terms were computed with the software Pajek64
Source(s): Created by the authors, using results from study

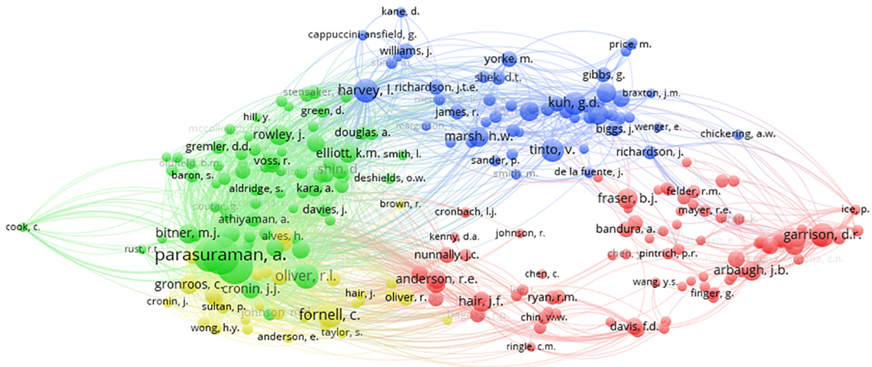


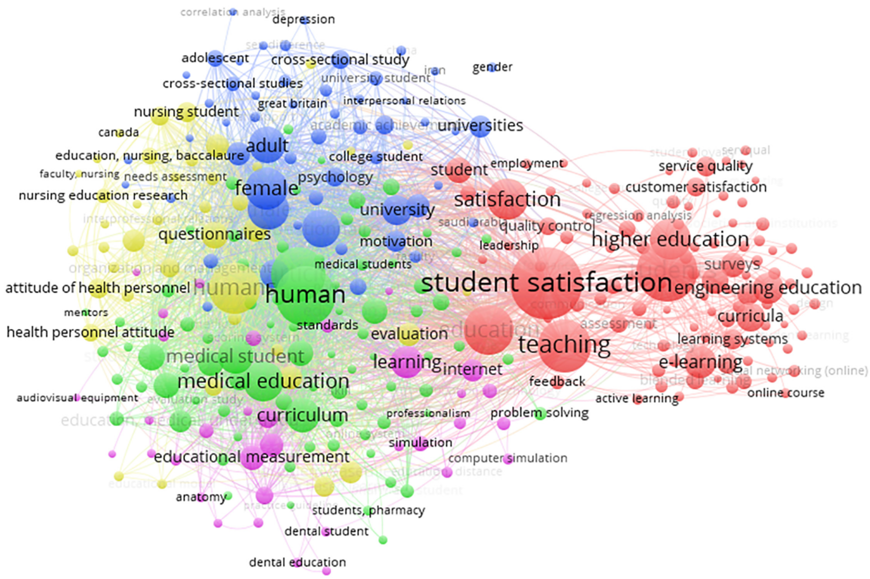
Figure 1.
Network map of co-citation analysis for authors (2002–2016)

Source(s): Created by authors, using results from the study. Threshold = 25 for the minimum number of citations of an author

Findings on student retention through the analysed sample record
Within the sample (n = 1,962 records), researchers identified two kinds of conceptual models regarding student persistence: data-driven and theoretically driven.

In the data-driven group, we find theories that posit student persistence is an outcome of personal and contextual characteristics: on the personal aspect, we find variables such as age, sex, socioeconomic status, academic performance or previous education, while in the contextual characteristics, the variables are teacher-related factors (e.g. attitude towards students, availability, experience or qualifications) and campus-related factors (e.g. equipment and installations, facilities or services). In this conceptual model, most the of studies use different forms of regression analysis, such as logistic regression, multilevel modelling or multiple linear regression (Dewitz *et al.*, 2009; Dey andAstin, 1993; Murtaugh *et al.*, 1999; Soria *et al.*, 2014).

In the theoretically driven group, we found conceptual frameworks developed in consolidated disciplines, such as marketing or psychology. Examples of the theories



Source(s): Created by authors, using results from the study. Threshold = 10 for the minimum number of occurrences of a keyword

Figure 2.
Network map of co-word analysis for keywords (2002–2016)

Table 3.
Clusters of co-cited authors specialized in students' retention (2002–2016)

Cluster	Colour	Author
1	Red	W. C. Black, J. F. Hair, R. E. Anderson, J. C. Nunnaly, J. Cohen, B. J. Babin, R. P. Bagozzi, K. Swan, D. F. Larcker, A. Bandura, S. Lee, P. Shea, Y. Chen, Y.S. Wang
2	Green	A. Parasuraman, R. L. Oliver, V. A. Zeithaml, J. Davies, L. L. Berry, D. Shin, J. Rowley, E. Kaynak, A. Douglas, J. Douglas, J. J. Cronin, M. Joseph, A. Kara
3	Blue	H. W. Marsh, P. Ramdsen, D. Kember, V. Tinto, G.D. Kug, A. W. Astin, R. James, L. Harvey, K. Wilson, S. Brown, A. Lizzio, J, Williams, J. T. E. Richardson
4	Yellow	R. L. Tatham, C. Fornell, M. D. Johnson, R. Brown, S. Taylor, T. Anderson, M. Raposo, A. Gustafsson, Y. Yi, P.M. Bentler, H. Alves, E. W. Anderson

Source(s): Created by the authors, using results from study

originated from marketing are applications of the SERVQUAL (Al-Alak and Alnaser, 2012; Emanuel and Adams, 2006; Mansori *et al.*, 2014) model to student satisfaction in an attempt to consider students and parents as clients of a firm, despite the nature of colleges and universities (public or private). In the case of theories originated from psychology, we can see derivations of the theory of planned behaviour focussing on the behavioural intentions of satisfied students with provided services or Bandura's theory of self-efficacy (DeWitz and Walsh, 2002). Within this framework our research has shown that researchers have a preference to test hypotheses using the current variants of structural equation modelling (SEM), such as covariance-based (CB) or partial least squares (PLS), given the complex structure of data involved for the analysis, and that in some studies it is possible to work with first- or second-order latent constructs (Copeland and Levesque, 2010; Kerby, 2015; Napoli and Wortman, 1998; Nora, 1987; Torres, 2006).

In the midst of both perspectives, there are some approximations that bridge the different factors influencing student retention into a robust conceptual model (Tinto, 1987, 2006). For example, a very well-known study on this topic (Cabrera *et al.*, 1993) explored how to integrate the two main theoretical frameworks for understanding student retention: student integration model (Tinto, 1975, 1982) and student attrition model (Bean, 1983, 1985). Researchers worked with a CB-SEM perspective and obtained an integrated model with appropriate goodness-of-fit indicators, strong path coefficients and explained variances higher than 40% for student persistence (a proxy variable of student retention).

We consider that both approaches are valuable for the theory development on student retention. In the initial stages, there are no robust conceptual models for the examined subject matter (e.g. students from different geographical regions or students from developing countries). Therefore, in these scenarios the best approximation is to use the available data.

Discussion

Visual representations under the bibliometric approach have the potential to make student retention research less complex to understand. There is a constant criticism that academic research produces results that are not relevant to decision-making processes; however, the use of visual representations allows decision-makers to better understand how the relevant issues have been addressed. It is important to add that this is a tailored conceptual approach, and the type of visual representation depends on the search strategy used when the topic was explored. Thus, the bibliometric approach generates personalized representations that avoid ambiguity and depend on the context posed by the decision-makers. This approach is consistent with Shang and Wang (2018), who identified key issues in emerging areas such as green supply chain management.

Student retention is a research field with more than 30 years of history. Throughout this time, different theoretical models have been proposed and validated. One of the more solid theories was developed by Vincent Tinto, whose model still worked as the reference paradigm during the first years of the last decade. However, his centrality has diminished in the last five years, when new robust conceptual models have been consolidated; for example, Napoli's and Wortman's model of psychological factors and Torres' retention model for Latino students at urban commuter universities have gained recognition among scholars specialized in student retention at higher education institutions.

This change in authors' centrality goes in parallel with an increased interest in conceptual models supported by advanced statistical approaches, such as multilevel modelling or SEM. This is interesting if we take into account that by the end of the 1990s, conceptual models were mainly tested by first-generation statistical techniques such as multiple linear or logistic regression analysis.

Regarding theoretical contributions, we observed that marketing scholars, such as A. Parasuraman, Valery A. Zeithaml or Leonard L. Berry, have consolidated the research front on student retention after 2007, because from 2002 to 2006 the influence of marketing scholars was not yet clear, as expressed by their betweenness centrality scores (see Table A1 in the Appendix 1). However, in the last 12 years, the research agenda focussed on service quality and customer satisfaction and consolidated the research undertaken by marketing scholars who considered the high costs of attracting new customers (i.e. students) to higher education institutions. In that sense, marketing scholars recognized that the mechanisms for student retention included a tailor-made service for them, focussed on improving their academic performance, social adjustment, as well as timely responses to problematic issues and demands, in order to secure student loyalty and identification with their *alma mater*.

In regard to the methods and research designs, we found a cluster with well-known authors of publications on methodology: Barry J. Babin, Joseph F. Hair, Rolf E. Anderson and

William C. Black for multivariate data analysis; Albert Bandura and Jum C. Nunnally for research designs and Jacob Cohen for effect size indicators and power analysis. The presence of this cluster is a sign that future research on student retention will introduce robust conceptual models tested with advanced multivariate techniques. This would result in the consolidation of this research front.

As we have seen, analytic techniques from bibliometrics (e.g. co-citation or co-word analysis, centrality measures) can be very helpful for identifying patterns within the academic literature. However, it should be used cautiously because there is always a risk of obtaining quick high-impact visualizations and colourful graphics. For this reason, it is advisable to include a specialist or subject expert, so we have an in-depth understanding of the analysed topic.

The adoption of the information visualization approach for accelerating and improving decision-making is an understudied area. Despite of this, results reported in academic literature show different criteria and best practices (e.g. adjusting by levels of tax complexity or different user perceptual types, as well as metrics to judge the quality of patterns within visualizations) to make the most of its benefits for increasing the quality of the decisions (Behrisch *et al.*, 2018; Teets *et al.*, 2010; Zabukovec; Jaklič, 2015).

In that sense, the study has an integrated perspective to improve our understanding of the subject, by reducing the various facets of each topic to focus intuitively on the most important. However, we must keep in mind that it is a process that requires the support of a specialist to contextualize the visual representations. Thus, the decision-making process on issues related to higher education becomes more efficient and has better support based on empirical data.

This study introduces an integrated perspective to better understand the topic and foster decision-making processes by making them more evident (visually) and expedite.

Replicability of the study

One of the most important things in a study is the possibility to replicate it to consolidate it in routine business processes. In this case, these are decision-making processes in higher education centres, which increase student satisfaction providing consequent positive repercussions, for both the institutions and the students.

The starting point of the study was basically the search strategy. Therefore, in order for it to be replicable, the researcher leading a new study should use almost the same search strategy for each period, to obtain similar results, after following all the steps described in the Method section. If some of the parameters were to change, the results may vary.

Data cleaning is a second consideration to take into account. Due to time and resource limitations, we were not able to standardize the authors' names (e.g. J. F. Hair vs J. Hair) or term variants (e.g. problem based learning vs problem-based-learning.). For this reason, some centrality scores could be above or below their real value in the processed data. Since we worked with all records retrieved from the multidisciplinary database, this issue should have affected just a few cases. Nevertheless, it is advisable to standardize authors' names and term variants before beginning the data analysis to assure precision throughout the study.

The time interval is the last issue to bear in mind before replicating our study. We decided to work from 2002 to 2017 when a critical mass of studies was indexed in databases, and we divided it in five-year periods, as explained before, due to the life cycle of the citations and the stability and reliability of measures of citation impact based on this temporal segmentation. However, if a future researcher were to determine a different starting point and period fragmentation, then the results may vary. Once there is a minimum bare of context-specific studies, it is advisable to test theoretically driven frameworks, because those studies require researchers to establish causal relationships based on previous findings.

Conclusions

We used 1,962 student retention studies to determine whether these could be analysed using bibliometrics to create a visual rendering of the results for a faster and more effective decision-making process. We are on the Internet and big data era where great amounts of data are created very fast and could lead to uncertainty (Koski, 2000; Voss, 2000). Therefore, we need to adopt an efficient approach for understanding the knowledge flow in critical areas, in this particular case, student satisfaction and retention, for our decision-making capability and our role as effective leaders in the learning and education environment.

We have provided the rationale behind the parameters of the study, the tools and its limitations. Consequently, this study can be applied to the same subject matter or another, where decision-making processes are needed; for example, strategies for improving learning experiences at MBA programmes, pedagogical approaches for fostering entrepreneurial behaviour among postgraduate students or appropriate learning styles for integrating interpersonal skills into business schools programmes.

We have provided evidence that a bibliometric approach in conjunction with data visualization can be a valuable complement to in-depth literature reviews for a decision-making process. We recommend the use of the bibliometric approach for building a rapid literature review or as an immersion mechanism for a discipline or research front.

The bibliometric approach used in this study let us identify clusters of authors specialized in student retention and detect the primary themes within this research field. We conclude that centrality measures provide an effective analytic strategy to: (1) examine the structure within a research front and (2) understand how authors can be organized as members of different research communities.

The managers of higher education centres must make many decisions every day, but they do not always have the necessary and relevant information, and often decide based on their previous experience. Decision-makers will appreciate the potential of bibliometric techniques because it contributes to making decisions aimed at improving the organizational management, to the extent that the complexity and diversity of problems can be reduced using easy and intuitive visual representations.

References

Aksnes, D.W. and Sivertsen, G. (2019), "A criteria-based assessment of the coverage of scopus and web of science", *Journal of Data and Information Science*, Vol. 4 No. 1, pp. 1-21, doi: 10.2478/jdis-2019-0001.

Al-Alak, B.A. and Alnaser, A.S.M. (2012), "Assessing the relationship between higher education service quality dimensions and student satisfaction", *Australian Journal of Basic and Applied Sciences*, Vol. 6, pp. 156-164.

Al-Kassab, J., Ouertani, Z.M., Schiuma, G. and Neely, A. (2014), "Information visualization to support management decisions", *International Journal of Information Technology and Decision Making*, Vol. 13, pp. 407-428, doi: 10.1142/S0219622014500497.

Baldeon Egas, P.F., Gaibor Saltos, M.A. and Toasa, R. (2020), "Application of data mining and data visualization in strategic management data at Israel Technological University of Ecuator", *ICAETT 2019 Advances in Intelligent Systems and Computing*, Vol. 1066, Springer, Cham, pp. 419-431.

Bean, J.P. (1983), "The application of a model of turnover in work organizations to the student attrition process", *The Review of Higher Education*, Vol. 6, pp. 129-148, doi: 10.1353/rhe.1983.0026.

Bean, J.P. (1985), "Interaction effects based on class level in an explanatory model of college student dropout syndrome", *American Educational Research Journal*, Vol. 22, pp. 35-64, doi: 10.3102/00028312022001035.

Behrisch, M., Blumenschein, M., Kim, N.W., Shao, L., El–Assady, M., Fuchs, J., Seebacher, D., Diehl, A., Brandes, U., Pfister, H., Schreck, T., Weiskopf, D. and Keim, D.A. (2018), "Quality metrics for information visualization", *Computer Graphics Forum*, Vol. 37, pp. 625-662, doi: 10.1111/cgf.13446.

Braun, J., Koch, C., Davis, J.L. and Miller, G.A. (1956), "The magical number seven, plus or minus two: some limits on our capacity for processing information", *Psychological Review*, Vol. 63 No. 2, pp. 81-97, doi: 10.1037/h0043158.

Cabrera, A.F., Nora, A. and Castañeda, M.B. (1993), "College persistence: structural equations modeling test of an integrated model of student retention", *Journal of Higher Education*, Vol. 64, pp. 123-139, doi: 10.2307/2960026.

Campanario, J.M. (2011), "Empirical study of journal impact factors obtained using the classical two-year citation window versus a five-year citation window", *Scientometrics*, Vol. 87, pp. 189-204, doi: 10.1007/s11192-010-0334-1.

Chen, H.M. (2019), "Information visualization skills for academic librarians: a content analysis of publications and online libguides in the digital humanities", *Library Hi Tech*, Vol. 37, pp. 591-603, doi: 10.1108/LHT-01-2018-0012.

Cho, J. (2014), "Intellectual structure of the institutional repository field: a co-word analysis", *Journal of Information Science*, Vol. 40, pp. 386-397, doi: 10.1177/0165551514524686.

Cios, K.J. and Kurgan, L.A. (2005), "Trends in data mining and knowledge discovery", in Nikhil, P. (Ed.), *Advanced Techniques in Knowledge Discovery and Data Mining*, Springer-Verlag, London, pp. 1-26.

Cios, K.J., Swiniarski, R.W., Pedrycz, W. and Kurgan, L.A. (2007), *Data Mining: A Knowledge Discovery Approach*, Springer, Boston, MA.

Claster, W., Shanmuganathan, S. and Ghotbi, N. (2008), "Text mining of medical records for radiodiagnostic decision-making", *Journal of Computers*, Vol. 3, pp. 1-6.

Copeland, K.J. and Levesque-Bristol, C. (2010), "The retention dilemma: effectively reaching the first-year university student", *Journal of College Student Retention: Research, Theory and Practice*, Vol. 12, pp. 485-515. doi: 10.2190/CS.12.4.f.

Daradkeh, M. and Abul-Huda, B. (2017), "Incorporating uncertainty into decision-making: an information visualisation approach", lecture notes in business information processing", *3rd International Conference on Decision Support System Technology*, ICDSST 2017, Namur, Belgium, 29-31 May 2017, Vol. 282, pp. 74-87, doi: 10.1007/978-3-319-57487-5_6.

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

DeWitz, S.J. and Walsh, W.B. (2002), "Self-efficacy and college student satisfaction", *Journal of Career Assessment*, Vol. 10, pp. 315-326, doi: 10.1177/10672702010003003.

DeWitz, S.J., Woolsey, M.L. and Walsh, W.B. (2009), "College student retention: an exploration of the relationship between self-efficacy beliefs and purpose in life among college students", *Journal of College Student Development*, Vol. 50, pp. 19-34. doi: 10.1353/csd.0.0049.

Dey, E.L. and Astin, A.W. (1993), "Statistical alternatives for studying college student retention: a comparative analysis of logit, probit, and linear regression", *Research in Higher Education*, Vol. 34, pp. 569-581, doi: 10.1007/BF00991920.

Dorta-González, P. and Dorta-González, M.I. (2013), "Impact maturity times and citation time windows: the 2-year maximum journal impact factor", *Journal of Informetrics*, Vol. 7, pp. 593-602, doi: 10.1016/j.joi.2013.03.005.

Emanuel, R. and Adams, J.N. (2006), "Assessing college student perceptions of instructor customer service via the Quality of Instructor Service to Students (QISS) Questionnaire", *Assessment and Evaluation in Higher Education*, Vol. 31, pp. 535-549, doi: 10.1080/02602930600679548.

Frantzi, K., Ananiadou, S. and Mima, H. (2000), "Automatic recognition of multi-word terms: the C-value/NC-value method", *International Journal on Digital Libraries*, Vol. 3, pp. 115-130, doi: 10.1007/s007999900023.

Gerdes, H. and Mallinckrodt, B. (1994), "Emotional, social, and academic adjustment of college students: a longitudinal study of retention", *Journal of Conseling and Development*, Vol. 72, pp. 281-288, doi: 10.1002/j.1556-6676.1994.tb00935.x.

Hoelscher, J. and Mortimer, A. (2018), "Using Tableau to visualize data and drive decision-making", *Journal of Accounting Education*, Vol. 44, pp. 49-59, doi: 10.1016/j.jaccedu.2018.05.002.

Ittoo, A.R., Zhang, Y. and Jiao, J.R. (2006), "A text mining-based recommendation system for customer decision making in online product customization" *IEEE International Conference on Management of Innovation and Technology*, ICMIT 2006, Singapore, 21-23 June 2006, pp. 473-477, doi: 10.1109/ICMIT.2006.262208.

Jacsó, P. (2009), "Five-year impact factor data in the journal citation reports", *Online Information Review*, Vol. 33 No. 3, pp. 603-614, doi: 10.1108/14684520910969989.

Jacsó, P. (2010), "Differences in the rank position of journals by Eigenfactor metrics and the five-year impact factor in the Journal Citation Reports and the Eigenfactor Project web site", *Online Information Review*, Vol. 34 No. 3, pp. 496-508, doi: 10.1108/14684521011054099.

Kerby, M.B. (2015), "Toward a new predictive model of student retention in higher education: an application of classical sociological theory", *Journal of College Student Retention: Research, Theory and Practice*, Vol. 17, pp. 138-161, doi: 10.1177/1521025115578229.

Koski, J.T. (2000), "Reflections on information glut and other issues in knowledge productivity", *Futures*, Vol. 33, pp. 483-495. doi: 10.1016/S0016-3287(00)00092-6.

Liu, X.Z. and Fang, H. (2020), "A comparison among citation-based journal indicators and their relative changes with time", *Journal of Informetrics*, Vol. 14, p. 101007, doi: 10.1016/j.joi.2020.101007.

Lu, H.-M., Chen, H., Chen, T.-J., Hung, M.-W. and Li, S.-H. (2010), "Financial text mining: supporting decision making using web 2.0 content", *IEEE Intelligent Systems*, Vol. 25, pp. 78-82.

Mansori, S., Vaz, A. and Ismail, Z.M.M. (2014), "Service quality, satisfaction and student loyalty in Malaysian private education", *Asian Social Science*, Vol. 10, pp. 57-66, doi: 10.5539/ass.v10n7p57.

Marshakova-Shaikevich, I. (1973), "System of document connections based on references", Nauchno-Tekhnicheskaya Informatsiya Seriya 2—Informatsionnye Protsessy i Sistemy

, Vol. 6, pp. 3-8.

McCullogh, R. (2019), "CiteScore 2018 metrics now available", available at https://blog.scopus.com/posts/citescore-2018-metrics-now-available (accessed 1 January 2020).

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

Moro, S., Cortez, P. and Rita, P. (2015), "Business intelligence in banking: a literature analysis from 2002 to 2013 using text mining and latent Dirichlet allocation", *Expert System with Applications*, Vol. 42, pp. 1314-1324, doi: 10.1016/j.eswa.2014.09.024.

Moya-Anegón, F., Chinchilla-Rodríguez, Z., Vargas-Quesada, B., Corera-Álvarez, E., Muñoz-Fernández, F., González-Molina, A. and Herrero-Solana, V. (2007), "Coverage analysis of Scopus: a journal metric approach", *Scientometrics*, Vol. 73, pp. 53-78, doi: 10.1007/s11192-007-1681-4.

Murtaugh, P.A., Burns, L.D. and Schuster, J. (1999), "Predicting the retention of university students", *Research in Higher Education*, Vol. 40, pp. 355-371, doi: 10.1023/A:1018755201899.

Nagda, B.A., Gregerman, S.R., Jonides, J., Von Hippel, W. and Lerner, J.S. (1998), "Undergraduate student-faculty research partnerships affect student retention", *The Review of Higher Education*, Vol. 22, pp. 55-72, doi: 10.1353/rhe.1998.0016.

Napoli, A.R. and Wortman, P.M. (1998), "Psychosocial factors related to retention and early departure of two-year community college students", *Research in Higher Education*, Vol. 39, pp. 419-455, doi: 10.1023/A:1018789320129.

Nora, A. (1987), "Determinants of retention among Chicano college students: a structural model", *Research in Higher Education*, Vol. 26, pp. 31-59, doi: 10.1007/BF00991932.

Oe, H., Yamaoka, Y. and Hideshima, E. (2016), "How to support decision making of local government in prioritising policy menu responding to citizens' views: an exploratory study of text mining approach to attain cognitive map based on citizen survey data", *Lecture Notes in Business Information Processing*, Vol. 250, pp. 202-216, doi: 10.1007/978-3-319-32877-5_15.

Pinto, M. (2015), "Viewing and exploring the subject area of information literacy assessment in higher education (2000-2011)", *Scientometrics*, Vol. 102, pp. 227-245, doi: 10.1007/s11192-014-1440-2.

Poreau, B. (2015), "Mapping South African public health research", *South African Medical Journal*, Vol. 105, pp. 52-55, doi: 10.7196/SAMJ.8866.

Quattrone, P. (2017), "Embracing ambiguity in management controls and decision-making processes: on how to design data visualisations to prompt wise judgement", *Accounting and Business Research*, Vol. 47 No. 5, pp. 588-612, doi: 10.1080/00014788.2017.1320842.

Rodrigues, S.P., van Eck, N.J., Waltman, L. and Jansen, F.W. (2014), "Mapping patient safety: a large-scale literature review using bibliometric visualisation techniques", *BMJ Open*, Vol. 4, p. e004468, doi: 10.1136/bmjopen-2013-004468.

Rubtsova, Y. (2014), "Automatic term extraction for sentiment classification of dynamically updated text collections into three classes", *Communications in Computer and Information Science*, Vol. 468, pp. 140-149, doi: 10.1007/978-3-319-11716-4_12.

Shang, W. and Wang, J. (2018), "Mapping the landscape and evolutions of green supply management", *Sustainability*, Vol. 10 No. 3, p. 597, doi: 10.3390/su10030597.

Small, H.G. (1973), "Co-citation in the scientific literature: a new measure of the relationship between two documents", *Journal of the American Association for Information Science*, Vol. 24, pp. 265-269, doi: 10.1002/asi.4630240406.

Soria, K.M., Fransen, J. and Nackerud, S. (2014), "Stacks, serials, search engines, and students' success: first-year undergraduate students' library use, academic achievement, and retention", *The Journal of Academic Librarianship*, Vol. 40, pp. 84-91, doi: 10.1016/j.acalib.2013.12.002.

Surjandari, I., Dhini, A., Wibisana, N. and Lumbantobing, E.W.I. (2015), "University research theme mapping: a co-word analysis of scientific publications", *International Journal of Technology*, Vol. 6, pp. 410-421, doi: 10.14716/ijtech.v6i3.1462.

Teets, J.M., Tegarden, D.P. and Russell, R.S. (2010), "Using cognitive fit theory to evaluate the effectiveness of information visualizations: an example of using quality assurance data", *IEEE Transactions on Visualization and Computer Graphics*, Vol. 16, pp. 841-853, doi: 10.1109/TVCG.2010.21.

Thomas, L. (2002), "Student retention in higher education: the role of institutional habitus", *Journal of Education Policy*, Vol. 17, pp. 423-442, doi: 10.1080/02680930210140257.

Tinto, V. (1975), "Dropout from higher education: a theoretical synthesis of recent research", *Review of Educational Research*, Vol. 45, pp. 89-125, doi: 10.3102/00346543045001089.

Tinto, V. (1982), "Limits of theory and practice in student attrition", *Journal of Higher Education,* Vol. 53, pp. 687-700, doi: 10.1080/00221546.1982.11780504.

Tinto, V. (1987), "From theory to action: exploring the institutional conditions for student retention", in Smart, J.C. (Ed.), *Higher Education: Handbook of Theory and Research*, pp. 51-89, University of Chicago Press, Chicago, IL, doi: 10.1007/978-90-481-8598-6_2.

Tinto, V. (2006), "Research and practice of student retention: what next?", *Journal of College Student Retention: Research, Theory and Practice*, Vol. 8, pp. 1-19, doi: 10.2190/4YNU-4TMB-22DJ-AN4W.

Torres, V. (2006), "A mixed method study testing data-model fit of a retention model for latino/a students at urban universities", *Journal of College Student Development*, Vol. 47 No. 3, pp. 299-318, doi: 10.1353/csd.2006.0037.

Vera-Baceta, M., Thelwall, M. and Kousha, K. (2019), "Web of science and scopus language coverage", *Scientometrics*, Vol. 121, pp. 1803-1812, doi: 10.1007/s11192-019-03264-z.

Verbene, S., Sappelli, M., Hiemstra, D. and Kraaij, W. (2016), "Evaluation and analysis of term scoring methods for term extraction", *Information Retrieval Journal*, Vol. 19, pp. 510-545, doi: 10.1007/s10791-016-9286-2.

Vílchez-Román, C. and Alhuay-Quispe, J. (2016), "Use of text mining for understanding Peruvian students and faculties' perceptions on bibliometrics training CEUR Workshop Proceedings", *3rd Annual International Symposium on Information Management and Big Data*, SIMBig 2016, Cusco, Perú, 1-3 September 2016, Vol. 1743, pp. 165-169, available at: http://ceur-ws.org/Vol-1743/paper20.pdf.

Vílchez-Román, C. and Quiliano-Terreros, R. (2017), "Stability and change in public studies in Colombia, Mexico and Peru: an exploratory approach based on co-word analysis", *Pan American Journal of Public Health*, Vol. 42, p. e35, doi: 10.26633/RPSP.2018.35.

Vílchez-Román, C., Huamán-Delgado, F. and Sanguinetti-Cordero, S. (2019), "Topic modeling applied to business research: a Latent Dirichlet Allocation (LDA)-based classification for organization studies", *Communications in Computer and Information Science,* Vol. 898, pp. 212-291, doi: 10.1007/978-3-030-11680-4_21.

Voss, D. (2000), "Scientists weave new-style webs to tame the information glut", *Science*, Vol. 289, pp. 2250-2251.

Waltman, L. (2016), "A review of the literature on citation impact indicators", *Journal of Informetrics*, Vol. 10, pp. 365-391, doi: 10.1016/j.joi.2016.02.007.

Wetzstein, A., Feisel, E., Hartmann, E. and Benton, W.C. (2019), "Uncovering the supplier selection knowledge structure: a systematic citation network analysis from 1991 to 2017", *Journal of Purchasing and Supply Management*, Vol. 25 No. 4, p. 100519, doi: 10.1016/j.pursup.2018.10.002.

Zabukovec, A. and Jaklič, J. (2015), "The impact of information visualisation on the quality of information in business decision-making", *International Journal of Technology and Human Interaction*, Vol. 11, pp. 61-79, doi: 10.4018/ijthi.2015040104.

Zentner, A., Covitt, R. and Guevara, D. (2019), "Exploring effective data visualization strategies in higher education", *SSRN Electronic Journal*, doi: 10.2139/ssrn.3322856, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3322856.

Zhu, D., Wang, D., Hassan, S.-U. and Haddawy, P. (2013), "Small-world phenomenon of keywords network based on complex network", *Scientometrics*, Vol. 97, pp. 435-442, doi: 10.1007/s11192-013-1019-3.

Appendix 1. Centrality measures for co-cited authors in each period

Rank	Author	Closeness	Betweenness
1	Tinto, V.	0.5512	0.0937
2	Marsh, H.W.	0.5412	0.0497
3	Tatham, R.L.	0.5371	0.0476
4	Astin, A.W.	0.5357	0.0535
5	Swan, K.	0.5330	0.0291
6	Shea, P.	0.5290	0.0218
7	Pickett, A.	0.5237	0.0182
8	Fredericksen, E.	0.5198	0.0157
9	Parasuraman, A.	0.5085	0.0198
10	Terenzini, P.T.	0.5072	0.0169

Note(s): Centrality values for authors were computed with the software Pajek64
Source(s): Created by the authors, using results from study

Table A1.
Ranking of authors with higher centrality values (2002–2006)

Rank	Author	Closeness	Betweenness
1	Parasuraman, A.	0.7341	0.0174
2	Zeithaml, V.A.	0.7222	0.0140
3	Ramsden, P.	0.7206	0.0301
4	Berry, L.L.	0.7059	0.0113
5	Rowley, J.	0.7011	0.0085
6	Tinto, V.	0.6996	0.0214
7	Oliver, R.L.	0.6918	0.0105
8	Fornell, C.	0.6842	0.0158
9	Shin, D.	0.6812	0.0063
10	Harvey, L.	0.6812	0.0068

Note(s): Centrality values for authors were computed with the software Pajek64
Source(s): Created by the authors, using results from study

Table A2.
Ranking of authors with higher centrality values (2007–2011)

Rank	Author	Closeness	Betweenness
1	Anderson, R.E.	0.9028	0.0093
2	Parasuraman, A.	0.8862	0.0062
3	Black, W.C.	0.8834	0.0086
4	Zeithaml, V.A.	0.8807	0.0056
5	Hair, J.F.	0.8780	0.0086
6	Berry, L.L.	0.8754	0.0054
7	Oliver, R.L.	0.8701	0.0052
8	Fornell, C.	0.8521	0.0053
9	Douglas, J.	0.8521	0.0051
10	Shin, D.	0.8471	0.0052

Note(s): Centrality values for authors were computed with the software Pajek64
Source(s): Created by the authors, using results from study

Table A3.
Ranking of authors with higher centrality values (2012–2016)

Appendix 2. Text mining of keywords in each period

Rank	Term	*C*-value	Closeness	Betweenness
1	Student satisfaction	70.300	0.8078	0.0567
2	Medical student	46.000	0.7669	0.0140
3	Medical education	35.667	0.7072	0.0087
4	Education programme	26.000	0.7467	0.0124
5	Higher education	18.000	0.5791	0.0047
6	Programme evaluation	17.780	0.6879	0.0081
7	Psychological aspect	17.000	0.6837	0.0056
8	Student attitude	13.000	0.6102	0.0021
9	Customer satisfaction	12.526		
10	Distance education	11.462		0.0014

Note(s): Centrality values for terms were computed with the software Pajek64
Source(s): Created by the authors, using results from study

Table A4.
Multi-word terms with higher centrality values (2002–2006)

Rank	Term	*C*-value	Closeness	Betweenness
1	Student satisfaction	276.974	0.9911	0.0753
2	Medical education	90.000	0.7708	0.0112
3	Higher education	76.182	0.6588	0.0146
4	Medical student	56.000	0.7115	0.0051
5	Programme evaluation	31.600	0.7279	0.0082
6	Customer satisfaction	30.791		
7	Educational measurement	28.000	0.6727	0.0029
8	Student attitude	27.000	0.6768	0.0037
9	Psychological aspect	27.000	0.6768	0.0032
10	Education programme	26.000	0.6748	0.0034

Note(s): Centrality values for terms were computed with the software Pajek64
Source(s): Created by the authors, using results from study

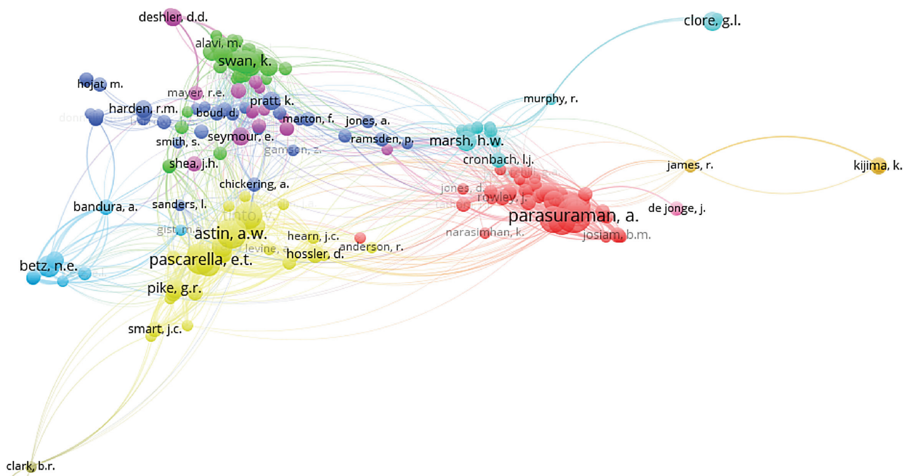
Table A5.
Multi-word terms with higher centrality values (2007–2011)

Rank	Term	*C*-value	Closeness	Betweenness
1	Student satisfaction	494.705	0.9865	0.0346
2	Medical education	162.636	0.8391	0.0080
3	Medical student	127.667	0.7993	0.0059
4	Higher education	115.414	0.6738	0.0104
5	Service quality	64.605		0.0014
6	Problem-based learning	57.059	0.7065	0.0038
7	Personal satisfaction	57.000	0.7739	0.0049
8	Educational measurement	56.000	0.7631	0.0047
9	Programme evaluation	52.500	0.7276	0.0033
10	Online learning	47.476		

Source(s): Created by the authors, using results from study
Note(s): Centrality values for terms were computed with the software Pajek64

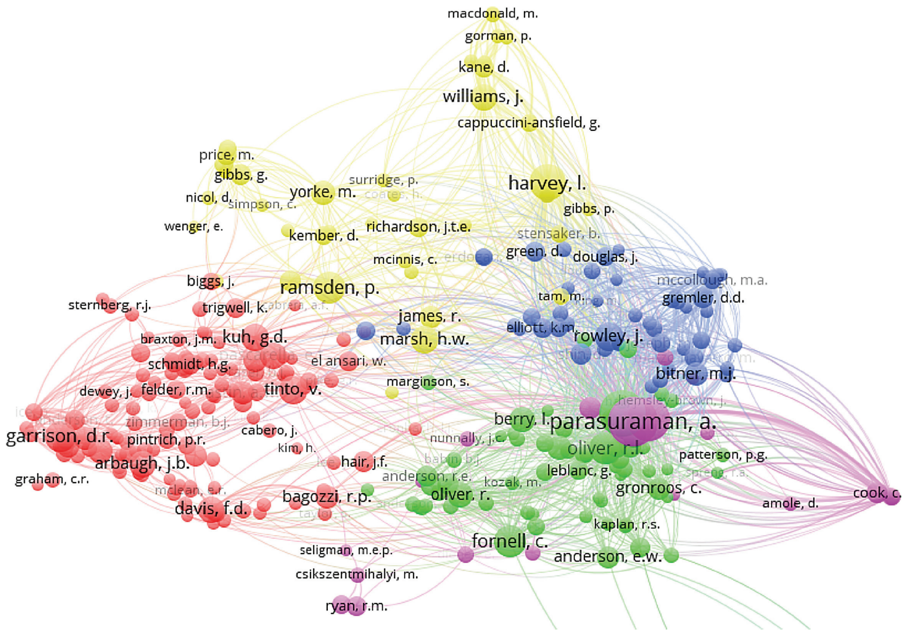
Table A6.
Multi-word terms with higher centrality values (2012–2016)

Appendix 3. Visualization of clusters and authors/keywords in each period



Source(s): Created by authors, using results from study

Figure A1.
Network map of co-citation analysis for authors (2002–2006)



Source(s): Created by authors, using results from study
Authors M.Lannario and D. Piccolo were excluded from the map to improve visualization

Figure A2.
Network map of co-citation analysis for authors (2007–2011)

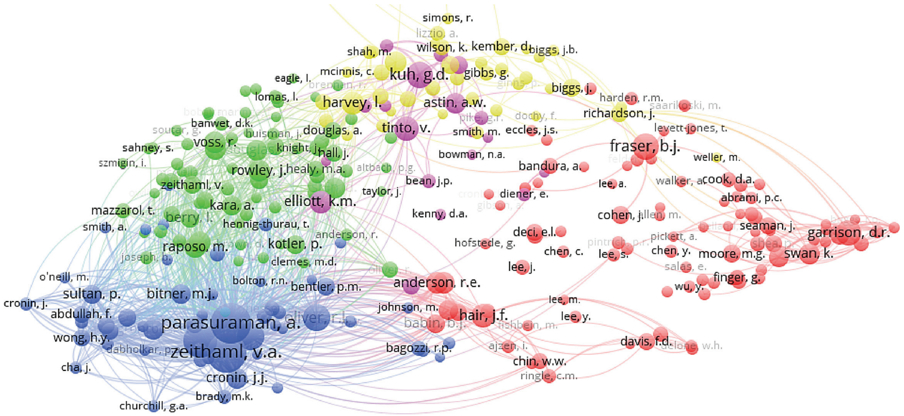


Figure A3.
Network map of co-citation analysis for authors (2012–2016)

Source(s): Created by authors, using results from study
Authors R.F. Lusch, D.T. Shek and R. C. Sun were excluded from the map to improve visualization

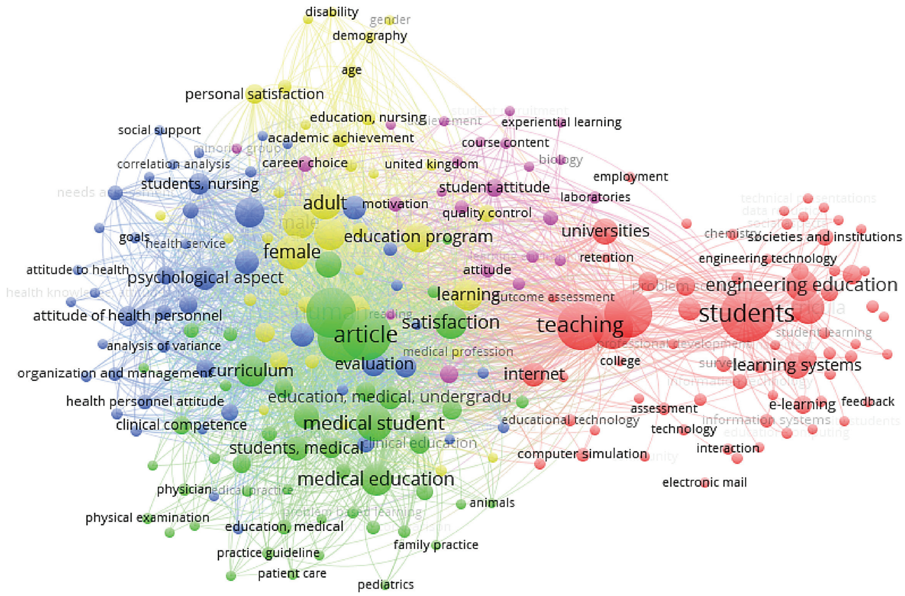
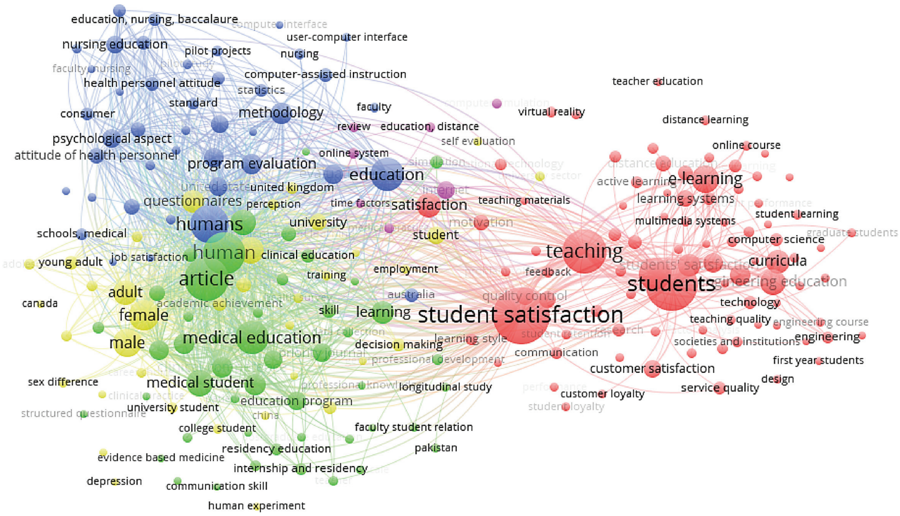


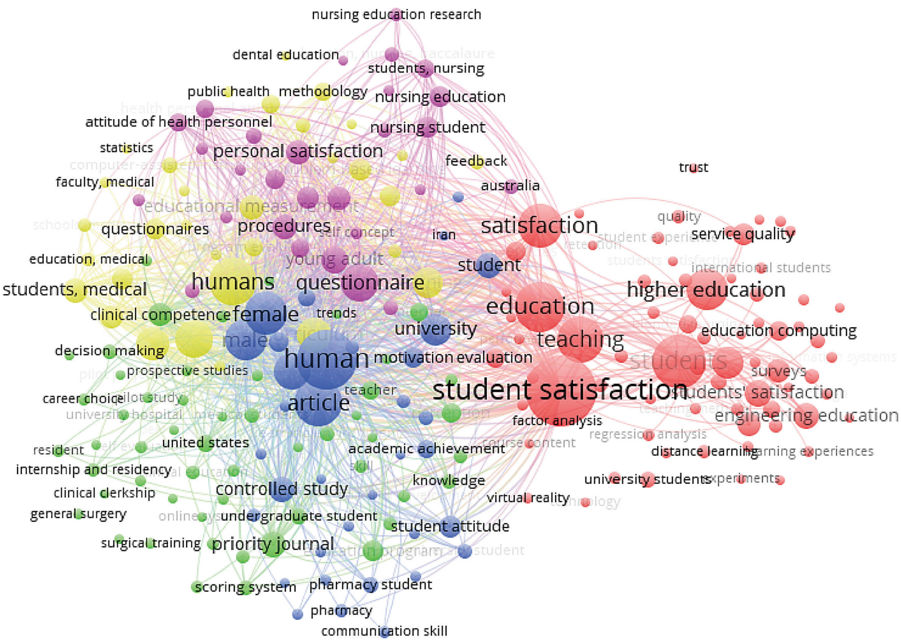
Figure A4.
Network map of co-word analysis for keywords (2002–2006)

Source(s): Created by authors, using results from study



Source(s): Created by authors, using results from study

Figure A5.
Network map of co-word analysis for keywords (2007–2011)



Source(s): Created by authors, using results from study

Figure A6.
Network map of co-word analysis for keywords (2012–2016)

Corresponding author
Carlos Vílchez-Román can be contacted at: cvilchez@pucp.edu.pe