ARTICLE

Is Assurance of Learning going as it should? A text mining-based review of literature

Andrés Chiappe [a]
Alejandro Emilio Wills [b]
Rolando Andrés Roncancio Rachid [c]

Abstract

Guaranteeing quality in Higher Education in today's globalized and uncertain context strains universities worldwide. Assurance of Learning (AoL) and Curricular Alignment (CA) are some of intimately related key concepts embedded in this situation. Albeit being pervasive in today's complex Higher Education ecosystem, AoL is not understood the same way everywhere. In order to explore its related meanings a scholarly literature analysis was carried out using text-mining techniques on 576,180 terms of 167 peer reviewed articles. Results reveals that research on AoL/CA is focusing on accreditation-oriented operationalization of assessment rather than synergically synchronizing the Intended Learning Outcomes (ILOs) with teaching and assessment as it should be. Evidence shows that AACSB accreditation guidelines have the heaviest impact on AoL/CA research which entails that a single point of view is globally driving what should otherwise be a dialogue in a diverse international academic ecosystem.

Keywords: Assurance of Learning. Curricular alignment. Text Mining Analysis. Higher Education Quality. AACSB.

1 Introduction

The issue of quality in Higher Education is as old as universities are. As early as the thirteenth century, two basic criteria for quality of university Education could be clearly distinguished in Western Europe: external accountability to higher authorities –ecclesiastical at first, governmental later – and internal peer review, a

[a] Universidad de La Sabana, Chía, Colombia.

[b] Universidad de La Sabana, Chía, Colombia.

[c] Universidad de La Sabana, Chía, Colombia.

mechanism that sought both the assurance of quality and autonomy through self-regulation (CRAFT, 2018). Along with universities, these early structural elements still survive today as active manifestations of the social purpose of universities.

From a social standpoint, one may contend that the most significant transformation in Higher Education along these centuries lies in shifting from an essentially elite system to what we now may consider as a massive provision of highly qualified professionals (HARVEY; WILLIAMS, 2010; JACKSON; BOHRER, 2010). As we come close to turning the first quarter of the twenty-first century, the expansion of the Higher Education system in the current globalized context of a technology-based economy has consequently made quality in Higher Education an increasingly enquired topic.

Such an increasing expansion of Higher Education offering is one of today's determinant issues regarding its quality. There are limitations worldwide in the available resources to guarantee educational quality, a problem aggravated by an interconnected context with ever-higher mobility of graduates between countries (CRAFT, 2018). The latter resulted in the development of both local and international organisms devised to guarantee the conditions for Higher Education. These organizations have created what can be considered de facto educational standards in the form of guidelines and other requirements for accreditation. Adoption of such requirements appears to be a matter of prestige, aimed at increasing student recruitment in a world where "competition between institutions for student enrolments has grown intense over the first decade of the century and becomes more critical as availability of places shifts to exceed demand" (SZEKERES, 2010). These guidelines also seem to become the main available mechanism for quality assurance in Higher Education across diverse fields. Such guidelines have become internationally recognized solutions to a "need for greater accountability to accompany the significant increase in the size of the student population and the students' greater expectations" (LOMAS, 2007).

Although universities do have an intellectual raison to be related to search, enhance, and expand humankind's knowledge, it must be considered that academic and business worlds intersect, and consequently that Higher Education also serves a more basic and practical societal need: to develop a workforce by providing highly-skilled, employable graduates to cover the needs of production systems (BATES, 2015; MA; VIEIRA, 2021; VOORHEES; HARVEY, 2005). As a consequence, the expansion of Higher Education as a service is unlikely to refrain. Moreover, a framework for understanding Higher Education quality should also include measurement or evaluation of their graduates' skills and attributes. They are in

fact required to demonstrate their competence and suitability for their subsequent professional performance (BECKER *et al.,* 2018; TRELEAVEN; VOOLA, 2008).

In order to cope with the challenges of expansion, and having to respond to an ever-increasing demand, universities – like everything else in the world – have turned towards the adoption of information technologies. But as technology enters the scene, it changes the scene (BATES, 2015). Whilst earlier analyses indicated a "peripheral and limited" use of Information and Communication Technology (ICT) in Higher Education (SELWYN, 2007), phenomena like the 'MOOC revolution' has thrust rates of adoption exponentially (VANDEYAR, 2020). Congruently, the traditional models of Higher Education are being challenged by diverse new models in constant evolution. As technology adoption in Higher Education has increased, more flexible and personalized competencies-based programs have emerged. Along with them comes the need for quality assurance (JOHNSON *et al.*, 2016). At this point, it is evident that instrumentalization of quality in Higher Education is rising (JARVIS, 2014). It is a consequence of the need to assure adequate levels in the provision of educational services under the current conditions and challenges.

As part of such efforts, the twin concepts of Assurance of Learning (AoL) and Curricular Alignment (CA) have emerged. CA is a broad principle seeking to align intended learning objectives – also known as content standards or curriculum standards – with instructional activities, supporting materials, and assessments; the latter includes standardized tests (ALLEN, 2004; ANDERSON, 2002). Thus, AoL "refers to processes for demonstrating that students achieve learning expectations for the programs in which they participate." (AACSB INTERNATIONAL, 2018, p. 32). The AoL process can be understood as an ongoing cycle of continuous improvement in educational programs from an assessment of student learning, in the following terms:

Step 1: Determine learning goals.

Step 2: Determine through teaching, where and how goals are addressed.

Step 3: Specify direct measures to assess learning goals.

Step 4: Establish performance outcomes and assess learning results.

Step 5: Close the loop (LAFLEUR; BABIN; LOPEZ, 2009).

As a matter of fact, research interest on CA and AoL has grown along with the development of educational technologies. According to Scopus' results, the research on AoL has increased constantly during the last 30 years, as shown in Figure 1.

Figure 1 - Number of Scopus-referenced articles meeting search criteria by year
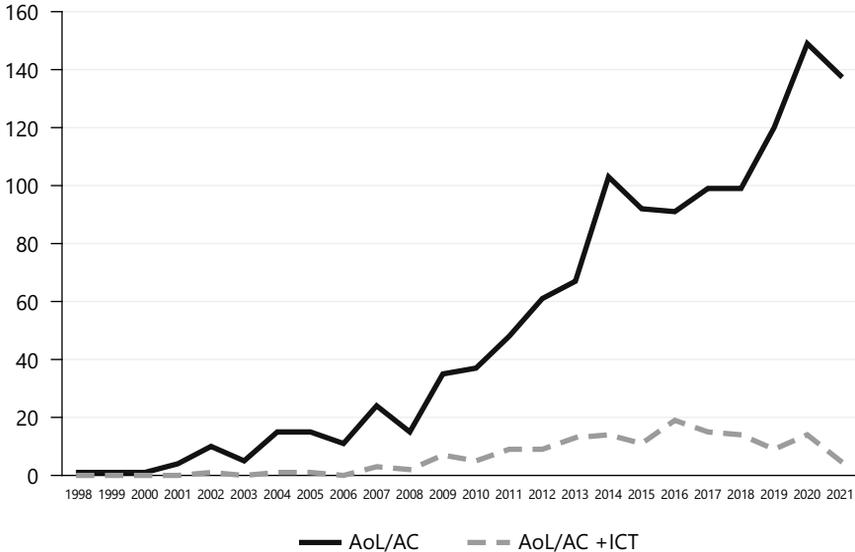
Source: Own elaboration based on Scopus data, 2022

2   Methodology

Considering the above ideas regarding quality in Higher Education, and in order to explore the elements involved, as well as to gain a better understanding of the current state of research on AoL/CA, a literature review on the subject was proposed using text mining techniques.

2.1   Text Mining and its goals

Following Miner *et al*., "Text mining and text analytics are broad umbrella terms describing a range of technologies for analyzing and processing semi-structured and unstructured text data." (2012, p. 30). This is, by definition "a highly interdisciplinary research field utilizing techniques from computer science, linguistics, and statistics" (FEINERER; HORNIK, 2008, p. 19).

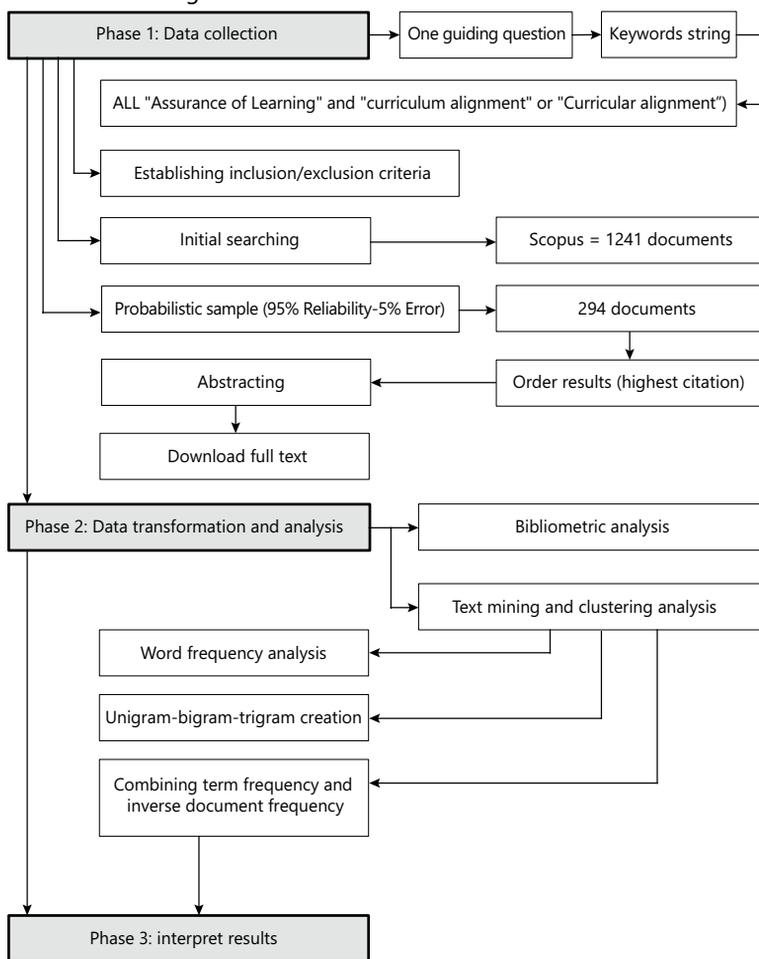As in many research tasks, the scholarly reviewer faces the tension between qualitative and quantitative. The difficulty in defining a text mining review lies in that the process seeks to extract qualitative results from quantitative, statistical observation of the language used in scholarly literature. Based on the former statement, for this review it was decided to use this approach to explore

the available literature in order to discover new facts hidden in plain sight on researchers' language.

2.2 Process breakdown

The scholarly literature review based on text mining followed a practical approach mainly based on the works of Feinerer and Hornik. (2008) and Welbers, Van Atteveldt and Benoit. (2017). In order to explain the process, the steps are presented as follows on Figure 2.

Figure 2 - Method Diagram

```
┌──────────────────────────────┐     ┌──────────────────┐     ┌──────────────────┐
│ Phase 1: Data collection     │────▶│ One guiding      │────▶│ Keywords string  │
│                              │     │ question         │     │                  │
└──────────────────────────────┘     └──────────────────┘     └──────────────────┘

     ┌──────────────────────────────────────────────────────────────────────┐
     │ ALL "Assurance of Learning" and "curriculum alignment" or             │
     │ "Curricular alignment")                                               │
     └──────────────────────────────────────────────────────────────────────┘

     ┌──────────────────────────────────────────┐
     │ Establishing inclusion/exclusion criteria │
     └──────────────────────────────────────────┘

     ┌──────────────────────────┐     ┌──────────────────────────┐
     │ Initial searching        │────▶│ Scopus = 1241 documents  │
     └──────────────────────────┘     └──────────────────────────┘

     ┌──────────────────────────────────────────┐     ┌──────────────────────────┐
     │ Probabilistic sample (95% Reliability-5%  │     │ 294 documents            │
     │ Error)                                    │     │                          │
     └──────────────────────────────────────────┘     └──────────────────────────┘

     ┌──────────────────────────┐     ┌──────────────────────────────┐
     │ Abstracting              │◀────│ Order results (highest       │
     │                          │     │ citation)                    │
     └──────────────────────────┘     └──────────────────────────────┘

     ┌──────────────────────────┐
     │ Download full text       │
     └──────────────────────────┘

┌──────────────────────────────────────────┐     ┌──────────────────────────────┐
│ Phase 2: Data transformation and analysis │────▶│ Bibliometric analysis        │
└──────────────────────────────────────────┘     └──────────────────────────────┘

                                                  ┌──────────────────────────────┐
                                                  │ Text mining and clustering   │
                                                  │ analysis                     │
                                                  └──────────────────────────────┘

     ┌──────────────────────────────┐
     │ Word frequency analysis      │
     └──────────────────────────────┘

     ┌──────────────────────────────┐
     │ Unigram-bigram-trigram       │
     │ creation                     │
     └──────────────────────────────┘

     ┌──────────────────────────────┐
     │ Combining term frequency and │
     │ inverse document frequency   │
     └──────────────────────────────┘

┌──────────────────────────────┐
│ Phase 3: interpret results   │
└──────────────────────────────┘
```

Source: Own elaboration, 2022

2.3 Data Collection

The first step for beginning with the data collection process was to set a purposeful research question in accordance with the scope of the scholarly literature review. The review question was: "Which are the main AoL topics that educational researchers are currently conducting studies on?"

In order to retrieve a sufficient but focused text corpus, the next searching string was applied to Scopus search engine: "Assurance of Learning" OR "curriculum alignment" OR "curricular alignment". The initial search yielded 1241 documents published between 1986 and 2021. Also, to configure a more suitable set of documents to analyze, a probabilistic 95% reliability-5% error sample (n = 294) was chosen.

Then, the sample was listed for full-text download from the Scopus database. Afterwards, they were ranked in descending order by number of citations, and their abstracts were checked for topic relevance as inclusion/exclusion criteria. Their research results were also verified for pertinency, and the final results were used in the analysis (n = 167).

2.4 Data transformation and analysis

Up to this point, the review process differed very few from any other reviews in existence. But then, once having identified the main topics and issues at hand from the most cited papers, the whole corpus was subject to text mining techniques.

To provide a better understanding and characterization of the analyzed sample of articles, they were subjected to retrieving general information about the text data sample, checking for more relevant sources and keywords using bibliometric techniques, and carrying out an exploratory data analysis based on word frequencies.

For this review, the R statistical computing platform was used because it offers a practical approach to text mining. It hosts a series of specialized applications (called *packages*) provided by an ecosystem of data analysts from all areas of science.

2.4.1 Bibliometric analysis

In order to determine the general impact on the scientific community and the levels of scientific activity for the sampled articles, this analysis followed the

techniques devised by Aria and Cuccurullo (2017). The bibliometric data was retrieved from a query on the online Scopus database.

2.4.2 Text mining analysis

The review team collected a very specific corpus of scholarly literature to sample a specific knowledge domain and to analyze its language. This computational linguistic analysis is not to be considered as mere automation, but as a new way to approach the reviews of scholarly literature. This analysis can be compared to an x-ray image that tells the researcher what to check and look for.

It seems necessary to identify the technical terminology, as well as the basic processes involved in the review. First, we carried out a frequency analysis. The text analysis units were called 'tokens', which can be *unigrams* (words), *bigrams* (pairs of words), or *trigrams* (triplets of words), in a 'bag-of-words' language model (ARRUDA; COSTA; AMANCIO, 2016). Once the initial frequency analysis was done, there was still a need to verify how informative the obtained tokens were in the context of individual documents. For this purpose, there was a "mechanism for attenuating the effect of terms that occur too often in the collection to be meaningful for relevance determination" (MANNING; RAGHAVAN; SCHÜTZE, 2008, p. 118). It is a heuristic weighting function known as TF-IDF which ranks tokens by combining *term frequency* (TF) within each document and the *inverse document frequency* (IDF) of the term in the whole corpus. In this way, terms that are too frequent in the entire *corpus* are 'punished' and more informative terms bubble up.

Finally, a networking analysis of terms was carried out, following the techniques presented by Arnold (2015). An adjacency matrix for the terms was built from their calculated Pearson correlation coefficients. This allowed the generation of a term network graphic, which showed a clustering structure that reveals the features in the text corpus.

The frequency analysis process deliberately left out stemming or lemmatization of tokens, a usual procedure in text mining analysis, thus maintaining lexical variations of terms, like plurals. Stemming would have conflated most lexical variations into a radix term, a single lexical form from which derivatives may stem – it can be an incomplete word, like 'soci-' for society, social, sociological etc. –; meanwhile, lemmatization would conform word occurrences to fewer terms by using a semantically proper relational approach, thus reducing variation to a meaningful *lemma*, a dictionary-like entry (MANNING; RAGHAVAN; SCHÜTZE, 2008).

The decision to refrain from these practices was made on the basis of communicational criteria from previous experiences, where readers found themselves guessing the original meaning of a stemmed word. Although researchers like Murakami *et al*. (2017) asserted that the effect of stemming requires further investigation, others had concluded that stemming "provides little measurable benefit and may in fact be harmful" in regard to semantic models like Latent Dirichlet Analysis, or LDA (SCHOFIELD; MIMNO, 2016, p. 298). The decision to waive stemming helped in maintaining as much information as possible, considering there is an inevitable loss of terms in the initial reading process, from hyphenated and misread words. At the time of this study, these are unsolved inherent software limitations. While most of the results came from frequency analysis, an LDA analysis was also carried out later by using the {topicmodels} R package (BLEI, 2012) in an attempt to directly address the research question.

2.5　Interpreting results

This is the last phase of the text data review, and it involves generating communicable knowledge from the results obtained. Knowledge rises from the interpretation of data retrieved during the text mining computational process, and from the reflective understanding of the process itself, i.e. there are also metacognitive results. On one side, significant patterns and relationships emerge as communicable insights from the data; on the other side, a flexible methodological pathway is taken in order to answer the research questions with consistent results.

3　Results

The results of the review from the application of the methodological process expressed in the previous figure are presented below. First, they are presented from the bibliometric perspective and then from the perspective of the review guiding questions.

3.1　Bibliometric results

The SCOPUS query was downloaded directly and produced the following results once fed into the R package. Table 1 shows data regarding bibliometric features of the analyzed articles and Table 2 presents data related with the quality of the top 10 consulted sources.

Table 1 - General bibliometric data

Documents	167
Sources (Journals, Books, etc.)	182
Period	1986 - 2021
Average citations per documents	8.007
Authors	691
Documents per Author	0.411
Authors per Document	2.43
Co-Authors per Documents	2.73
Collaboration Index	2.84

Source: Own elaboration, 2022

Table 2 - Top publishing sources by number of articles

Sources	Items	SJR Impact factor	SJR Quartile
Journal of Education for Business	44	0.508	Q2
Journal of Accounting Education	8	0.931	Q1
Advances in Accounting Education: Teaching and Curriculum Innovations	6	0.301	Q2
ASEE Annual Conference and Exposition Conference Proceedings	6	N/A	N/A
Journal of Marketing Education	5	0.976	Q1
Proceedings of ISECON	5	N/A	N/A
Journal of Chemical Education	4	0.499	Q1
Currents in Pharmacy Teaching and Learning	3	0.633	Q2
International Journal of Management Education	3	1.167	Q1
International Journal of Science Education	3	1.092	Q1

Source: Own elaboration, 2022

The most influential papers were identified by using the *h*-index (HIRSCH, 2005) of the whole collection according to Scopus, i.e. the *h* papers that received at least *h* citations, individually and excluding self-citations. The top 23 articles came from 14 journals, all of which are related to Education: 10 are dedicated to Social Sciences and Education (71.43%), 3 specialize in Business, Management and Accounting with an educational approach (21.43%), and one deals with Medical Education (7.14%). In order to discover other bibliometric features of the corpus, the {bibliometrix} R package was used to analyze the results of the Scopus query.

3.2   Exploratory Data Analysis

Initially, the articles were retrieved and preprocessed in two different ways to evaluate the consistency of results. Both analysis paths were carried out using R packages. The current state of the art in text analysis using R entails three main different structures for storing text data, and each one of them has its own processing packages. One of those paths entails the use of the {tm} R package to read the documents, clean the data and create a document-term matrix for analysis. Out of the various practical analysis tool sets available, this is mostly a one-stop approach (SWARTZ; SHRIVASTAVA, 2021), but not necessarily a straightforward one.

In another pathway, the {textreadr} and {tidytext} packages were used to create the required R data structures containing the individual elements of analysis. The process of obtaining a proper set of such elements is called tokenization. In both paths of analysis, tokenization of the entire corpus was carried out without term-conflating subprocedures of stemming or lemmatization (MANNING; RAGHAVAN; SCHÜTZE, 2008). The results were consistent with the previous ones, although the {tidytext} processing uses less computational resources, and the programming code is sleeker and clearer to read. This approach was preferred for obtaining the initial tokenization data; analysis using {tm} was later introduced for TF-IDF heuristics and LDA processing.

3.2.1   Frequency Analysis

The first results obtained from the analysis were individual term frequencies. The corpus was preprocessed to clean it from URLs, email addresses, numbers, punctuation signs, special characters, and extra white spaces. Then, English stop-words were removed and the text was reorganized in a one-row-per-term fashion (ALZOUBI, 2020). Once cleaned, the corpus totaled N = 576,180 terms, with n = 28,048 unique entries. The ratio N/n = 20.54 may be understood as a gross 'reuse' ratio, expressing how many times a unique term appeared in the corpus and it is equal to the mean of the calculated word frequencies, of course.

Before applying the TF-IDF weighting, document-term matrix sparsity was reduced. A high percentage of that matrix consists of zeroes from terms not found on particular documents. Removal of such sparse terms was done through the specific function provided in the {tm} package. A parameter of 0.8 was used, which means a j term with a document frequency $df_j > N*(1-0.8)$ will be retained, N being the number of documents. The resulting matrix has merely

1,718 terms, compared to 28,048 in the original count. That figure is consistent with a Pareto-like cut for the cumulative distribution function of frequencies for unigrams.

3.2.2 TF-IDF Weighting

The {tm} R-package implementation of the TF-IDF function was used on the reduced document-term matrix. The main result is an ordered list of highly relevant terms, whose top occurrences are shown on Table 3.

Table 3 - Top 20 relevant terms from the sparsity-reduced and TF-IDF weighted document-term matrix

Term (Unigram)	TF-IDF Weight	Frequency	Rank by Frequency
business	1.16	4397	7
accounting	0.89	6393	4
teachers	0.77	2409	14
aol	0.71	882	87
health	0.66	757	116
AACSB	0.53	1162	55
alignment	0.47	1573	36
teacher	0.45	1392	43
marketing	0.44	1074	62
scores	0.40	1247	49
medical	0.40	486	238
exam	0.35	954	76
mba	0.34	350	365
test	0.33	1729	31
faculty	0.31	2753	13
instructional	0.28	919	83
mathematics	0.28	523	207
technology	0.28	1163	54
score	0.27	678	134
ethics	0.27	711	125

Source: Own elaboration, 2022

In the first place, the TF-IDF unigram results reinforce the high relevance, as per ranking by the normalized distribution function, of "accounting" (2nd, 0.54), "AACSB" (6th, 0.33), "scores" (9th, 0.28), "test" (12th, 0.23) and "exam" (15th, 0.22). Two appearances that stand out come in the form of "health" (5th, 0.42) and "medical" (13th, 0.23) tokens. Reviewing the bibliometric results, there are 40 sampled journals classified in the Medicine (n = 13), Health Professions (n = 10), Psychology (n = 9), Nursing (n = 4), and Pharmacology, Toxicology and Pharmaceutics (n = 4). Their TF-IDF heuristic properly highlights their contribution. Relating to the trigram frequencies, "health" appeared in 669 entries; as for bigrams, it is part of 359 entries. "Medical" showed up in 369 trigrams and in 205 bigrams.

A remarkable token under TF-IDF ranking is "marketing" (7th, 0.32), that occupied a relatively lower position amongst non-weighted unigrams (54th, n = 1,020); "marketing" is part of 740 trigrams (the topmost ranked 101th) and of 438 bigrams (the topmost ranked 39th). One could say that literature production on marketing concerning AoL is lower than in other fields, as only 5 marketing-related journals out of 97 in the Business, Management and Accounting category were retrieved by the basic query. However, as a topic on individual articles, the TF-IDF function ranks it just below the AACSB.

An undoubtedly significant result of applying the TF-IDF heuristics is the appearance of "ethical" (17th, 0.20) and "ethics" (22nd, 0.18), otherwise ranked 221st and 107th in unweighted unigrams. The same goes for "behavioral" (19th, 0.20) and "mathematics (23rd, 0.18). Along the lines of didactical strategies, "competencies" ranks 26th (0.17), "instructional" is at the 28th position (0.17), "technology" ranks 33rd (0.16), and "textbooks" is 34th (0.16).

3.2.3   Unigrams, bigrams, and trigrams

As explained in the Methodological section above, tokens obtained in text processing can be either single words (unigrams), joined word couples (bigrams), or triplets (trigrams). The first results of inspecting bigram and trigram lists are the uninformative but expected outcomes, which are rapidly used as feedback for the cleanup filters. This, of course, means the process is recursive (in a human sense), taking the initial outputs to correct filters and produce better outputs iteratively. In a way, filters operate as a log of the process: in there, the names of academic institutions, main publishing houses, and their headquarters' cities show up. As those are put aside, a programmatic record of what was excluded remains in the code as a fringe result.

Once the cleaning process on data takes place, more unexpected but meaningful results start to appear. The first noticeable issue that emerged during cleaning and feedback preprocessing of unigrams was the need to separate the Association to Advance Collegiate Schools of Business (AACSB), acronym that repeatedly appeared attached to other words. A special function had to be programmed to complete the task. This observation alone highlights the pervasive appearances of the AACSB throughout the research literature on AoL that was sampled.

There is a series of statistical observations leading to consolidate the view of the AACSB as the preponderant factor in the reviewed academic production. The fact that the last update of AACSB guidelines was published in 2013 can be no coincidence. Then, the bibliometric results of local citations confirmed the organization's prevalence in literature, as shown. The team performed a simple check to determine on how many articles could the acronym itself be found, and how it ranked by frequency. It was present in 64 out of N = 163 articles, 39.26% of the corpus, and ranked in the 77th position among 28,048 unigrams. A total 566 trigrams and 372 bigrams also included the term 'AACSB'.

There are some expanded forms of the acronym to consider to further the analysis, as all these term appearances found in a sample of published research since 1986 speaks of the level of influence of the AACSB. The trigram "**advance collegiate schools**" was present in 62 articles (38%) and ranked 4th (n = 133) out of 100,618 tokens in the non-weighted trigram frequency; its n/N = 0.816 ratio means almost one mention per article for the whole corpus. The trigram "**aacsb accredited business**" ranked 55th (n = 25). On the bigram ranks, "**advance collegiate**" and "**collegiate schools**" stood in 34th and 35th place (n = 140 for both) among 145,558 total bigram entries, while "**aacsb accredited**" ranked 46th (n = 111).

By inspection of the trigrams, a somewhat expected but worth noting fact stands out. The highest-ranked trigram is "**accounting education literature**" (n = 187), and the word "**accounting**", which ranks fifth among all unigrams (n = 4,938), is mentioned in 68 articles of the sample (41.71%). It also appears in 840 bigrams and in 2387 trigrams. Related bigram "**accounting education**" also ranked 1st, followed by "**accounting educators**", which ranked 21st. As for trigrams, 18 tokens that include "**accounting**" ranked in the top 100. To verify the proportion of specialized sources, it is noted that 97 of the journals in the sample were classified in the "Business, Management and Accounting" category. Considering the many topics existing in the business management field, this

indicates either a preponderant interest among researchers, or simply a massive level of production from one source.

The next highlighted visual-inspection observation is the high rank of the "**major field test**" trigram, ranked 3rd in frequency (n = 137) and uniquely present in 18 articles (11%). Related trigram "**educational testing service**" ranked 15th (n = 55) and appeared in 24 articles (14.72%). Another related instance was "**ets major field**", ranked 20th in frequency (n = 44) and present in 13 (7.9%) out of 163 articles. While "**ets**" as unigram was mentioned 219 times in 20 articles, bigram "**major field**" is ranked 14th by frequency (n = 198), and mentioned in 25 articles. Bigram "**field test**" is ranked 22nd by frequency (n = 170), and mentioned in 21 articles. Finally, bigram "**ets major**" appeared 57 times in 16 articles out of 163 (9.0%), at an average 4.58 times per article, ranking in the 186th position by frequency. These observations show that researchers are considering ETS's examinations with interest. A loosely related trigram, "**multiple choice questions**" ranks 23rd (n = 40) and appears in 14 articles (8.58%). As a significant comparison, the trigram "**critical thinking skills**" ranks 26th, three positions below the latter.

One study that evaluates international students' and English teachers' perceptions on the TEFL MA Exam stood out, possibly indicating researcher's concern with internationalization of the student population. This refers to the appearance of the trigram "**tarjoman cogent social**", ranked 112th (n = 17), from the article titled "An Investigation of a Nationwide Exam from a Critical Language Testing Perspective" (SALEHI; TARJOMAN, 2017). Not only the study also fell within the sample and was hence *locally* cited, but it indicates what seems to be a negative backwash from the mentioned test.

3.2.4   Network Analysis on the TF-IDF Unigram Graph

As introduced in the methodology section and as shown un Figure 3, the team obtained a node-size and transparency-coded Fruchterman-Reingold network graph for the top-30 unigrams after TF-IDF weighting. It was plotted from an adjacency matrix of positive correlations using the {igraph} R package. Closer nodes are highly related; larger nodes have more positive correlations, and hence are more central and opaque.

Figure 3 - Node-size and transparency-coded Fruchterman-Reingold layout of a network graph for the top-30 TF-IDF unigrams
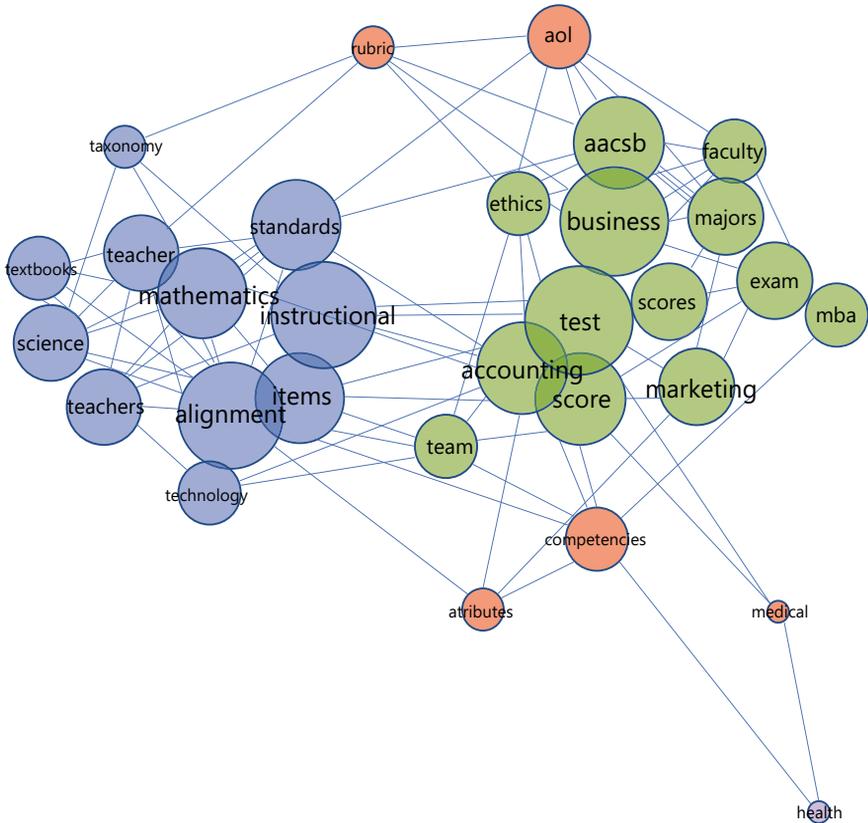
Source: Own elaboration, 2022

After deriving Pearson's correlations from the TF-IDF weighted Document-Term matrix, all positive results are kept as edges of a graph whose vertices or nodes are the top-connected terms. The Fruchterman-Reingold algorithm models the system as particles subject to forces of attraction (more connections) or repulsion (less connections) (ARNOLD, 2015). This automatically increases the centrality of "heavier" nodes. Here, two main clusters are identified. Although orientation of the graph occurs spontaneously on execution of its generative algorithm, this spatial distribution appears consistently: there is a heavier cluster where "business", "accounting" and "marketing" (as subject contents) interlace with "tests", "AACSB", "score/s", and "exam"; those nodes are surrounded by "majors", "faculty", and "MBA". Siding it, there is a less cramped cluster where "instructional", "alignment", and "standards" outweigh "mathematics", "science", "teacher/s", "textbooks" and "technology".

Looking at the degrees of connection (i.e. positive correlation) between most relevant terms, a first observation concerns the size and position of the main nodes of interest, namely "AoL" and "Alignment". On one hand, the "AoL" acronym for Assurance of Learning has five-to-one more connections to the first mentioned cluster. "AoL" is clearly related to the heavier group of concepts centered around "business" and surrounded by "test", "score/s", and "exam". Also in this cluster, and directly connected to "AoL", the "AACSB" node has a larger weight and a more central position. On the other hand, the key concept of "Alignment" (as part of "Curricular Alignment", or CA) is the heaviest central node from the other mentioned cluster. It is surrounded by two larger nodes, "instructional" and "standards", which suggests confirmation of an operationalized description of the CA concept.

The way this result can be interpreted here is that research is slightly polarized between "AoL" and "Curricular Alignment". The literature produced on AoL tends to be strongly "test-centric", as it gravitates around terms more related to test or exam-based assessment and AACSB as a particular accreditation institution.

4   Discussion

AoL is both a concept and a process. As a concept, it emerges from a need to guarantee educational quality, assuring learning outcomes, while as a process it was set up to produce constant improvement of educational programs. However, the evidence herein indicates that may end up endangering its very own purpose. While it is supposed to entail a connected awareness of objectives, teaching, contents and assessment, the results show an "eccentricity": there is a trend, visible through the results, for schools and researchers to lean upon a pragmatic application of assessment-based instructional programs in order to comply with accreditation requirements. Assessment would always be a good focal point for AoL processes, were it not that its immediate incarnation comes in the shape of score-focused standardized tests. AoL as a process must aim at building competencies and bring forth complex skill sets in business students. Evaluating such skills also requires a complex mechanism of assessment. In the very spirit of AoL, those mechanisms must be adequately aligned with the teaching practices, the contents and be part of the objectives. If universities keep finding that accreditation pressure overpowers educational objectives, it may be time to reevaluate the function of quality assurance and accreditation institutions.

It should be noted that the results presented here are statistical in character, but it is important to understand them as an X-ray image of the language used in

research. From numerical data emerges an image of concerns, ideas and thoughts that researchers are working on as they publish. One salient feature of this statistical approach is a reduced bias. The data is publicly available, as is the programming code (if needed), which means complete reproducibility of results. Also, any researcher can follow the same path and determine its consistency and robustness.

For social scientists, this apparently simple review entailed various challenges. One of them was appropriation of the computational tools required for developing a suitable result. This does not merely mean the aforementioned difficulty in acquiring or using the toolkits available; what is noteworthy is that there was a permanent tension between the weight of the methodological choices and the social research goals. Once established that computational analysis was the path to follow, there was a continuous adjusting between the momentum of computational dynamics in the information extraction process and the review goals. Extracting features from text data yields a lot of statistical information which needs to be properly framed for communication. An ancillary result of this tension is that purely technical and methodological issues become results in themselves, as shown.

As specialized as it may be, scientific literature is but a genre of writing. An analysis of the language employed in a particular sample of that kind of texts tells a lot about the way their scientific writers are thinking. This new perspective helps in discovering facts that could've gone unnoticed had this kind of text mining analysis not been performed. This exploratory analysis approach, profits from the very methodological issues we discovered while performing it, and the data results themselves change and determine the path of the technical processes applied.

From the point of view of the language used to talk about them, there is a qualitative difference between AoL and CA. The foremost evidence shows up as the two separate word clusters found in the Network Analysis section. CA is related more closely to teachers, textbooks and science than AoL, and that alone should trigger an alert on how the AoL practices are being construed by scholars, and hence eventually implemented by business schools.

However, since "teachers, textbooks and science" are terms also closely connected to Standards", which is in turn connected to the AACSB, further analysis is required. Following Anderson (2002), CA "standards" refer to Content Standards or Curriculum Standards, which equate to ***learning objectives***. On the other hand, as the interviewed Kathryn Martell clarifies regarding AACSB standards,

AoL is one of the sections therein, and it refers directly to **assessment of student learning**. She clearly states that AACSB has no intention of "standardizing **program assessment methods**", as the interviewer suggests is the direction of the AACSB guidelines.

The continuous program improvement purpose of student assessment is, in terms of the current analysis, <u>statistically underrepresented</u> in the research *corpus*. Indeed there is a semantic duality in the word "standards", as it is either applied to ILOs under CA, or understood as learning assessment under AoL; but beyond that, there is an observable lag in the frequencies of terms related to "improvement" or "program improvement".

As a metacognitive conclusion regarding the mentioned 'human recursiveness' of the process is that coding required for this kind of tasks cannot be a monolithic step-by-step application of procedural formulas. What actually sheds light on the path to follow is the reflective part of the 'design thinking' approach researchers have to use. It is an iterative process in which results themselves are the guide indicating what steps to take next, in a quest for obtaining meaningful insights with purposeful procedures. Evidence emerges at every step, which needs to be documented and acted upon. A final conclusion – which is also a caveat for future researchers – regarding current literature on computational text analysis and text data mining: for practical purposes, the relevant literature is focused on showing the practical 'how to' with workable examples. The eventual results from real data are messier than expected and perplexing at times.

Broadly speaking, the foregoing is perplexing insofar as it would be expected to be able to find more consistent approaches in the literature on matters of an evaluative and curricular nature. In this regard, the results of this review give way to reflect on this issue from at least two interesting perspectives: first, attending the current relevance or relevance of the "standard" concept in evaluation and second, from the perspective of CA, the prevalence of the closeness to the "standard of accreditation" and not to the consistency in the curricular design that ensures the quality of the educational service that is provided to the students.

In this sense, regarding the first matter, the assurance of the quality of Education through "standards" would reinforce the idea of a homogenizing Education in which the educational processes should lead to the fulfillment of the standards determined by the Ministry of Education for the students. The foregoing, although it constitutes a very frequent way of understanding the role and importance of educational standards, leaves aside an issue that is highly relevant for Education

in the 21st century: conducting learning processes that respond to the interests, expectations, limitations and students' personal learning paths (CHIAPPE *et al.,* 2020). This issue seems to be very relevant for future research processes that address the tensions between flexibility and standardization, at least from the point of view of evaluation. There would arise questions such as: should learning assessment mechanisms continue to monitor how well students have achieved learning standards? Or, on the contrary, should assessment be able to verify learning based on what each one have determined as your own learning objectives?

On the other hand, from the curricular perspective, it is possible to identify a risk of distortion of the very concept of "Curricular Coherence" if the management of the "standard" continues simply as the exercise of applying a norm or accreditation guideline. The foregoing is understood to the extent that it is recognized that learning experiences are very diverse in their scope, purpose and context and that the educational designs available to generate or support them must respond to said diversity. Thus, it is suggested to give the importance it deserves to the processes of ensuring consistency and coherence in the design and arrangement of the different components and processes that lead to learning, including the definition of learning objectives, the didactic proposal and resources and evaluation mechanisms.

A Garantia de Aprendizagem está indo como deveria? Uma revisão de literatura baseada em mineração de texto

Resumo

Garantir a qualidade do Ensino Superior no contexto globalizado e incerto de hoje gera tensão nas universidades de todo o mundo. Garantia de Aprendizagem (AoL) e Alinhamento Curricular (CA) são alguns dos conceitos-chave intimamente relacionados e incorporados nessa situação. Apesar de ser onipresente no complexo ecossistema de ensino superior de hoje, a AoL não é entendida da mesma maneira em todos os lugares. Para explorar seus significados relacionados, uma análise da literatura acadêmica foi realizada usando técnicas de mineração de texto computacional em 576.180 termos de 167 artigos revisados por pares. Os resultados revelam que a pesquisa da AoL/CA está focada, principalmente, na operacionalização da avaliação orientada para a acreditação, em vez de sincronizar, sinergicamente, os Resultados de Aprendizagem Esperados (RPA) com o ensino e a avaliação, como deveria ser. As evidências mostram que as diretrizes de acreditação da AACSB têm o maior impacto na pesquisa da AoL/CA, o que implica que um único ponto de vista está impulsionando globalmente o que deveria ser um diálogo em um ecossistema acadêmico internacional diversificado.

Palavras-chave: *Garantia de Aprendizagem. Alinhamento Curricular. Análise de Mineração de Texto. Qualidade do Ensino Superior. AACSB.*

¿La Garantía de Aprendizaje funciona como debería? Una revisión de la literatura basada en la exploración de textos

Resumen

Garantizar la calidad en la educación superior en el contexto globalizado e incierto de hoy genera tensión en las universidades de todo el mundo. La Garantía de Aprendizaje (AoL) y la Alineación Curricular (CA) son algunos de los conceptos clave íntimamente relacionados e incrustados en esta situación. A pesar de ser omnipresente en el complejo ecosistema de la educación superior actual, AoL no se entiende de la misma manera en todas partes. Con el fin de explorar sus significados relacionados, se llevó a cabo un análisis de la literatura académica utilizando técnicas de exploración de texto en 576.180 términos de 167 artículos revisados por pares. Los resultados revelan que la investigación sobre AoL/CA se centra principalmente en la operacionalización de la evaluación orientada a la acreditación en lugar de sincronizar sinérgicamente los Resultados Previstos de Aprendizaje (ILOs) con la enseñanza y la evaluación, como debería ser. La evidencia muestra que las pautas de acreditación de AACSB tienen el mayor impacto en la investigación de AoL/CA, lo que implica que un solo punto de vista está impulsando globalmente lo que de otro modo debería ser un diálogo en un ecosistema académico internacional diverso.

Palabras clave: *Aseguramiento del Aprendizaje. Alineación Curricular. Análisis de Exploración de Textos. Calidad de la Educación Superior. AACSB.*

References

AACSB INTERNATIONAL. 2*013 Eligibility procedures and accreditation standards for business accreditation*: engagement, innovation, impact. Tampa, 2018. Available from: https://www.aacsb.edu/-/media/aacsb/docs/accreditation/business/standards-and-tables/2018-business-standards.ashx?la=en&hash=B9A F18F3FA0DF19B352B605CBCE17959E32445D9. Access in: 2021 May 24.

ALLEN, M. J. *Assessing academic programs in higher education*. Bolton: Anker Publishing Co., 2004.

ALZOUBI, W. A. Dynamic graph based method for mining text data. *WSEAS Transactions on Systems and Control,* [s. l.], v. 15, p. 453-458, sep. 2020. https://doi.org/10.37394/23203.2020.15.45

ANDERSON, L. W. Curricular alignment: a re-examination. *Theory Into Practice*, [s. l.]. v. 41, n. 4, p. 255-260, Nov. 2002. https://doi.org/10.1207/s15430421tip4104_9

ARIA, M.; CUCCURULLO, C. *bibliometrix*: an R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, [s. l.], v. 11, n. 4, p. 959-975, nov. 2017. https://doi.org/10.1016/j.joi.2017.08.007

ARNOLD, T. *Humanities data in R*. New York: Springer Science+Business Media, 2015.

ARRUDA, H. F.; COSTA, L. F.; AMANCIO, D. R. Using complex networks for text classification: Discriminating informative and imaginative documents. *Europhysics Letters*, [s. l.], v. 113, n. 2, p. 28007, jan. 2016.

BATES, A. W. *Teaching in a digital age*: guidelines for designing teaching and learning for a digital age. British Columbia: BC Open Texbooks, 2015.

BECKER, S. A. *et al*. *The NMC horizon report*: 2018 higher education edition. Louisville: Educause, 2018.

BLEI, D. M. Probabilistic topic models. *Communications of the ACM*, [s. l.], v. 55, n. 4, p. 77-84, abr. 2012. https://doi.org/10.1145/2133806.2133826

CHIAPPE, A. *et al*. La educación del siglo XXI y el despertar de las bellas durmientes: una revisión sistemática de la literatura. *Education in the Knowledge Society (EKS)*, [s. l.], v. 21, n. 1, p. 1-15, 2020. https://doi.org/10.14201/eks.22483

CRAFT, A. *International developments in assuring quality in higher education*. New York: Taylor & Francis, 2018.

FEINERER, I.; HORNIK, K. Text mining infrastructure in R. *Journal of Statistical Software*, Innsbruck, v. 25, n. 5, p. 1-54, 2008. https://doi.org/10.18637/jss.v025.i05

HARVEY, L.; WILLIAMS, J. Fifteen years of quality in higher education. *Quality in Higher Education*, v. 16, n. 1, p. 3-36, Apr. 2010. https://doi.org/10.1080/13538321003679457

HIRSCH, J. E. An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, v. 102, n. 46, p. 16569-16572, Nov. 2005. https://doi.org/10.1073/pnas.050765510

JACKSON, S.; BOHRER, J. Quality assurance in higher education: recent developments in the United Kingdom. *Research in Comparative and International Education*, [s. l.], v. 5, n. 1, p. 77–87, mar. 2010.

JARVIS, D. S. L. Regulating higher education: quality assurance and neo-liberal managerialism in higher education: a critical introduction. *Policy and Society*, [s. .], v. 33, n. 3, p. 155-166, set. 2014. https://doi.org/10.1016/j.polsoc.2014.09.005

JOHNSON, L. *et al. NMC Horizon Report*: 2016 higher education edition. Austin: The New Media Consortium, 2016.

LAFLEUR, E. K.; BABIN, L. A.; LOPEZ, T. B. Assurance of learning for principles of marketing students: a longitudinal study of a course-embedded direct assessment. *Journal of Marketing Education*, [s. l.], v. 31, n. 2, p. 131-141, Apr. 2009. https://doi.org/10.1177/0273475309335242

LOMAS, L. Zen, motorcycle maintenance and quality in higher education. *Quality Assurance in Education*, Bingley, v. 15, n. 4, p. 402-412, Oct. 2007. https://doi.org/10.1108/09684880710829974

MA, J.; VIEIRA, D. M. Remando ou guiando o barco? Modelos de governança para as políticas de Educação Superior no Brasil pós-1990. *Ensaio: Avaliação e Políticas Públicas em Educação*, Rio de Janeiro, v. 29, n. 113, p. 1000-1023, Oct./Dec. 2021.

MANNING, C. D.; RAGHAVAN, P.; SCHÜTZE, H. *Introduction to information retrieval*. New York: Cambridge University Press, 2008.

MINER, G. *et al*. Practical text mining and statistical analysis for non-structured text data applications. Waltham: Academic Press, 2012.

MURAKAMI, A. *et al*. 'What is this corpus about?': using topic modelling to explore a specialised corpus. *Corpora*, Edinburgh, v. 12, n. 2, p. 243–277, Ago. 2017. https://doi.org/10.3366/cor.2017.0118

SALEHI, M.; TARJOMAN, M. An investigation of a nationwide exam from a critical language testing perspective. *Cogent Social Sciences*, [s. l.], v. 3, n. 1, p. 1396639, Jan. 2017. https://doi.org/10.1080/23311886.2017.1396639

SCHOFIELD, A.; MIMNO, D. Comparing Apples to Apple: the effects of stemmers on topic models. *Transactions of the Association for Computational Linguistics*, v. 4, p. 287-300, Dec. 2016. https://doi.org/10.1162/tacl_a_00099

SELWYN, N. The use of computer technology in university teaching and learning: a critical perspective: a critical look at computer use in higher education. *Journal of Computer Assisted Learning*, [s. l.], v. 23, n. 2, p. 83-94, Jan. 2007.

SWARTZ, S.; SHRIVASTAVA, A. Stepping up the game–meeting the needs of global business through virtual team projects. *Higher Education, Skills and Work-Based Learning*, [s. l.], v. 12, n. 2, p. 346-368, 2021. https://doi.org/10.1108/HESWBL-02-2021-0037

SZEKERES, J. Sustaining student numbers in the competitive marketplace. *Journal of Higher Education Policy and Management*, [s. l.], v. 32, n. 5, p. 429-439, Oct. 2010. https://doi.org/10.1080/1360080X.2010.511116

TRELEAVEN, L.; VOOLA, R. Integrating the development of graduate attributes through constructive alignment. *Journal of Marketing Education*, [s. l.], v. 30, n. 2, p. 160–173, Ago. 2008. https://doi.org/10.1177/0273475308319352

VANDEYAR, T. A window to teachers' ICT practices: Discerning between teaching and the complex science of pedagogy. *Ensaio: Avaliação e Políticas Públicas em Educação*, Rio de Janeiro, v. 28, n. 109, p. 982-1011, Oct./Dec. 2020. https://doi.org/10.1590/S0104-40362020002802388

VOORHEES, R. A.; HARVEY, L. Higher education and workforce development: a strategic role for institutional research. *New Directions for Institutional Research*, v. 2005, n. 128, p. 5-12, Dec. 2005. https://doi.org/10.1002/ir.159

WELBERS, K.; VAN ATTEVELDT, W.; BENOIT, K. Text analysis in R. *Communication Methods and Measures*, [s. l.], v. 11, n. 4, p. 245-265, Oct. 2017.

Information about the authors

Andrés Chiappe: PhD in Educational Sciences. Professor at the Universidad de la Sabana. Senior Researcher and peer evaluator of Minciencias for research programs and projects on topics related to the use of ICT in Education. Fellow of the Unesco/ICDE on the Open Educational Movement for Latin America and Founder/chairman of the Ibero-American Network of Researchers in Innovation and Educational Technology. Director of Doctorate in Educational Innovation using ICT. Contact: andres.chiappe@unisabana.edu.co

iD https://orcid.org/0000-0002-9664-4833

Alejandro Emilio Wills: Aspiring Master in Educational Informatics. Electrical Engineering, professional in literature and philosophy, Universidad de la Sabana. Contact: alejandro.wills@unisabana.edu.co

iD https://orcid.org/0000-0002-6307-7983

Rolando Andrés Roncancio Rachid: Master and Ph.D. in Government and Culture of Organizations, Master in Business Administration - MBA. Rector/Principal of the Universidad de la Sabana, Professor of Strategy and Corporate Governance at INALDE Business School and Senior Consultant for Organizational Transformation at Emergap. Contact: rolando.roncancio@unisabana.edu.co

iD https://orcid.org/0000-0002-4301-1178