Journal of Informetrics 11 (2017) 483–497

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Proposal of indicators for the structural analysis of scientific articles

José Osvaldo De Sordi[a,*], Wanderlei Lima de Paulo[b], Manuel Antonio Meireles[b], Marcia Carvalho de Azevedo[c], Luis Hernan Contreras Pinochet[c]

[a] Faculdades Metropolitanas Unidas (FMU), Rua Iwakuni, 236 Jundiaí – SP, 13211-424, Brazil
[b] Faculdade Campo Limpo Paulista (FACCAMP), Rua Guatemala, 167 Campo Limpo Paulista – SP, 13231-230, Brazil
[c] Universidade Federal de São Paulo (UNIFESP), Rua Angélica, 100 Osasco – SP, 06110-295, Brazil

ARTICLE INFO

Article history:
Received 28 October 2016
Received in revised form 28 February 2017
Accepted 28 February 2017
Available online 22 March 2017

Keywords:
Cross-reference
Article structure
Text matrix
Structural analysis
Article classification

ABSTRACT

This study aims to identify variables and indicators that substantiate the development of rules that focus on the structural analysis of scientific articles. Variables and indicators for structural analysis are derived from hypotheses deduced from editorials in important scientific journals. To exemplify and test the indicators, a structural analysis was conducted of 108 scientific articles published in important journals in the field of Management. The hypotheses were mostly tested in accordance with the idea of estimation statistics. The approach that was developed for the structural analysis of the network of texts innovates by employing network analysis indicators (indegree and outdegree). For this purpose, the text matrix is employed through the identification and encoding of cross-references between sections and subsections of each article under study. For the context in question, the field of Management, twelve rules were developed. The interpretations of the possible values for the indicators, expressed in the form of rules, are applied as directives to less experienced scholars in preparing their scientific articles, and for the generation of information to support activities concerning the classification and analysis of scientific articles.

1. Introduction

The importance of the structure of an article has been highlighted by the editors of prominent journals. Sun and Linton (2014, p. 571) wrote an editorial that pointed out that "paper writing is a critical step in publishing research work. Structure offers a basis, skeleton and acts as a guide – especially for multi-author collaborations". Bansal and Corley (2012), in an editorial, addressed the structural differences between the front end and back end of qualitative articles. This was one of seven editorials of the Academy of Management Journal, in which the editors gave suggestions and advice for improving the quality of articles to be submitted to the journal. The essence of these editorials is that there are common structural aspects among the articles published in high-impact journals that should be observed by researchers who intend to publish through these channels. These editorials, together with others such as that of Sparrowe and Mayer (2011), Zhang and Shaw (2012),

* Corresponding author.
E-mail addresses: osdesordi@gmail.com (J.O. De Sordi), wldepaulo@gmail.com (W.L. de Paulo), profmeireles@uol.com.br (M.A. Meireles), marciacarvalhoacevedo@gmail.com (M.C. de Azevedo), luisherman@terra.com.br (L.H.C. Pinochet).

http://dx.doi.org/10.1016/j.joi.2017.02.012
1751-1577/

invited us to reflect on the theme, especially on the possibility of identifying and defining discriminatory characteristics in scientific articles published in high-impact journals.

Considering that "a good editorial is an opinion maker" and "what it analyses can be the basis of the production of new evidence" (Singh & Singh, 2006; p. 15), we used the information from the cited editorials as a source of inspiration for the present study. The opinions found in editorials encouraged us to consider the structure of the scientific article, how it is divided into sections and the number of words, as commonly employed features. Some authors, such as Cargill and O'Connor (2009) and Sun and Linton (2014), make use of these features to represent graphically the predominant side view of the structure of scientific articles.

The focus and innovation of this study is to consider the structure of scientific articles in terms of the relationships between the sections of the article. Internal relationships are characterized by cross-references and other devices discussed in this study, which will be identified and encoded in text matrices for each of the articles in the sample. The analysis of the interrelations between sections will be based on indicators using network analysis techniques, applied to the context of textual documents, in accordance with the AnaCoTEx approach proposed by De Sordi, Meireles and De Oliveira (2016). The analyses of section size and the relationships between sections will include statistical tests to analyze the opinions of experienced researchers and editors, as declared in editorials and presented in this study in the form of hypotheses. As the analyses of the editors are mostly specific, according to the type of research, whether qualitative, quantitative or qualitative-quantitative (Creswell, 2003), the hypotheses and variables will be segmented by these types of research.

The high number of scientific articles rejected by prominent journals (Linton, 2012) and the importance of structural aspects (distribution of words, sections and subsections between front end and back end article texts, according to the demands of each type of research) for the quality assigned to the article (Sun & Linton, 2014), were among the principal motivations for this study. The study aims to identify variables and indicators that substantiate the development of rules that focus on the structural analysis of scientific articles. For this purpose, we analyzed 108 articles published in important journals in the field of Business Management. The knowledge derived from this study is of direct concern to a wide range of professionals involved in the development, analysis and classification of scientific articles.

2. Structure of scientific articles

"Empirical social science journal articles normally consist of six parts: (1) Introduction, (2) Literature review, (3) Methodology, (4) Result, (5) Discussion, and (6) Conclusion" (Sun & Linton, 2014; p.571). This standard structure was used in an analysis that considered only the volume of the sections (word count). For this study, which also addresses the interrelations between sections in terms of cross-references, we made two alterations to the standard structure described by Sun and Linton (2014): a) exclusion of the Conclusion section, as it has no association with cross-references (it does not cite, and is not cited by, the other sections) and is treated by many authors as a subsection of the Discussion; b) addition of the Appendix, as many articles have one or more appendices, resulting in many cross-references. These adaptations will be revisited and justified in Section 4, based on what was identified in the articles of the study sample. Thus, the standard structure of the scientific article considered in this study is composed of six sections: Introduction, Theory/Literature-review, Method, Results/Findings, Discussion and Appendix.

2.1. Size of sections

Sun and Linton (2014) used the number of words in the sections to conduct a comparative analysis of two groups of articles: 50 desk-rejected manuscripts recently submitted to Technovation and ten highly cited papers from Technovation. Bansal and Corley (2012) also worked with the idea of size, but rather than section, they worked on parts of the text, using the concepts of front end and back end of the articles, described and analyzed as follows:

> The front end of a quantitative article typically includes an introduction, literature review, and the development of new theory by way of hypotheses. The literature review, therefore, sets the background for the hypotheses. Because qualitative papers fulfill a different purpose, their front end is shorter, yet it serves more functions.
>
> [. . .]
>
> long, robust back end
>
> [. . .]
>
> Qualitative works, on the other hand, reserve the biggest punch for the back end. A strong Discussion section should not only summarize the findings and ultimately delineate the theoretical and practical implications that are also demanded of quantitative papers [. . .] (Bansal & Corley, 2012; p. 510).

The relationships highlighted in the editorial are "front end shorter" and "back end robust and long". Considering that we analyzed articles from different journals, encompassing different types of research, we will work on part of the analyses with the front end and back end concept. Like Bansal and Corley (2012), we will consider as the front end of the article all the sections that precede the Method section, with all the others being considered as the back end of the article. Thus, we will analyze the opinion of these authors using the following hypothesis:

H1–The ratio between the volume of words of the front end and the back end is a discriminatory characteristic of the type of research, whether qualitative, quantitative or qualitative-quantitative, being more equal (closer to one) for quantitative articles, more unequal (closer to zero) for qualitative articles and, for qualitative-quantitative articles, the result is a position between the two other types (closer to the middle).

2.2. Cross-references between sections

The Discussion section of qualitative articles is quite different from its counterpart in quantitative articles. Whereas the quantitative type tests what has been created and explained in the front end, in qualitative research, most of the epistemological efforts occur in the Discussion. In this section, the interpretations and creation of scientific knowledge take place, as highlighted by Bansal and Corley (2012, p.510):

> Qualitative works, on the other hand, reserve the biggest punch for the back end. A strong Discussion section should not only summarize the findings and ultimately delineate the theoretical and practical implications [. . .] but also integrate data and theory in a way that explicitly conveys the connections between the analyzed data, the emergent theory, and the literatures at which the contribution is aimed.

Bansal and Corley (2012) highlight a strong integration between data and theory in the Discussion section of qualitative articles. From this, it can be inferred that a large volume of cross-references to the other sections (outdegree) will be made in the Discussion section, leading to the following hypothesis:

H2–The analysis of cross-references of the Discussion section to the others (outdegree) is a discriminatory characteristic of the type of research, being higher in the case of qualitative studies.

Likewise, thinking of a large volume of cross-references that allow the types of research to be characterized, we sought for mentions and specific characteristics of articles of the quantitative type. Sparrowe and Mayer (2011) highlighted the importance of hypotheses for the construction of scientific knowledge in this type of research:

> Hypotheses are the heart of a paper, and grounding hypotheses is one of the most important tasks in crafting effective theory. [. . .] A great deal of thought goes into every paper, and the theory section is key to explaining how one is going to add value to the research topic and why these specific hypotheses make sense individually and fit together to form a coherent conceptual framework. (Sparrowe & Mayer, 2011; p. 1101).

What we infer from this editorial is that a large volume of citations to the section (indegree) where the hypotheses are declared (Literature review) can be expected, which leads to the formulation of the third hypothesis:

H3–The analysis of cross-references in the other sections to the Literature Review section (indegree) is a discriminatory characteristic of quantitative research, meaning that this section has the highest indegree of all the standard sections of the quantitative article.

3. Method

3.1. Selection of journals and articles for the sample

Our analysis included articles from widely cited journals in Business Management. The journals were selected in March 2015, using the following criteria: the journal had to be on the list of the 20 main journals in the field of Business Management in the Scopus (SCImago Journal Rank) and the Thomson Reuters (Journal Citation Reports – Social Sciences Citation Index) databases, and available for access, i.e., be accessible by the scientific article repositories contracted by our institutions. Nine journals matching these criteria were identified: Academy of Management Journal, Academy of Management Review, Journal of Consumer Research, Journal of Marketing, Journal of Marketing Research, MIS Quarterly, Organizational Research Methods, Organization Science, and Strategic Management Journal.

The following stage involved the identification of articles that declare the type of research as qualitative, quantitative or qualitative-quantitative, in accordance with the following criteria. Quantitative articles should include the words "quantitative" and "hypotheses" in the abstract or the term "quantitative research" in the abstract. Qualitative-quantitative articles should mention the words "quantitative" and "qualitative" in the abstract. For qualitative articles, three strategies of qualitative research were defined, and at least one of these should be described in the abstract: "grounded theory", "phenomenological" or "phenomenology", and "ethnography" or "ethnographic". Furthermore, for qualitative articles, these should not include the words "hypothesis", "hypotheses" or "quantitative" in the abstract. From these searches, we excluded texts classified as letters and research notes, annotated bibliography and articles that presented no applied research. Articles presented in media (files) that could not be edited were also excluded, as this is a requirement for the encoding of the articles to be read by the software that generates the text matrix (AnaCoTEx) to be used by the network analysis software and techniques. The application of these criteria resulted in 12 phenomenological articles, which defined the quantities of the rest of the samples of qualitative articles: 12 ethnographic articles and 12 grounded theory articles, resulting in 36 qualitative articles. Thus, another 36 quantitative articles and 36 qualitative-quantitative articles were selected at random, resulting in a sample of 108 articles for analysis. The cluster sampling of the qualitative, quantitative and qualitative-quantitative articles was done in a single stage. The types of research defined the clusters and, for each cluster, articles were selected

Table 1
Example of a text matrix generated by anacotex software to the article of Joy and Sherry (2003).

	1	2	3	4	4.1	4.2	4.3	4.4	4.5	4.6	4.7	5	Out Degree
1	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	1	0	0	0	0	0	0	0	0	1
4	0	0	0	0	0	0	1	1	1	1	1	0	5
4.1	0	0	0	0	0	0	0	0	0	0	0	0	0
4.2	0	0	0	0	2	0	0	0	0	0	0	0	2
4.3	0	0	0	0	4	0	0	0	0	0	0	0	4
4.4	0	0	0	0	7	0	0	0	0	0	0	0	7
4.5	0	0	0	0	3	0	0	0	0	0	0	0	3
4.6	0	0	0	0	3	0	0	0	0	0	0	0	3
4.7	0	0	0	0	4	0	0	0	0	0	0	0	4
5	0	0	0	0	0	0	0	0	0	0	0	0	0
In Degree	0	0	0	1	23	0	1	1	1	1	1	0	29

by convenience, considering the aforementioned criteria. Due to space restrictions for this text, the references of the 108 articles are provided in a supplementary file.

When we found that our repository of scientific articles included 34,335 documents from nine journals of interest, we conducted tests with some specific research terms for each of the three types of research, composing them in selection commands. The aim was to identify few, but pertinent, articles with only one of the three types of research. It was noted that articles in the field of Management are predominantly and historically quantitative, and that in this type of work the "hypotheses are the heart of a [business management] paper" (Sparrowe & Mayer, 2011; p. 1101). Therefore, the term "hypotheses" was used as a research criterion, combined with two other terms: "quantitative" and "quantitative research". In the case of qualitative articles, it was observed that these are fewer in number in the field of Management, although they are increasingly being used by researchers in the field (Bansal & Corley, 2012), as is the practice of triangulating strategies and research techniques. To select exclusively qualitative articles, without incorporating qualitative-quantitative ones, we developed a selection command made up of two sets of criteria: a) the first involving the names of typically qualitative research strategies (grounded theory, phenomenology, ethnography and words derived from these terms); b) the second, excluding articles linked to typically quantitative research terms. To identify articles associated with qualitative-quantitative research, the words qualitative and quantitative were used as research criteria. For the three selection processes, associated with the three types of research, we analyzed the articles resulting from the selection commands by skimming the articles to them (Duggan & Payne, 2009). This process enabled us to gauge the effectiveness of the criteria employed to identify articles using one and only one type of research.

3.2. Encoding of articles (identifying cross-references) for structural analysis

For the structural analysis of the relationships between the different sections and subsections of each of the 108 articles, the network analysis technique was used, being applied to the context of textual documents in accordance with the AnaCoTEx approach. In this approach, the sections, subsections and other subsequent levels of the article are considered as network actors. These actors constitute a square matrix known as the text matrix, which for network analysis constitutes a one-mode network, as all the actors are of the same type i.e., parts of text (section, subsection or sub-subsection). The relationships between the actors, described in the cells of the text matrix and formed by the intersection of different actors, is of the directional and valued type (Wasserman & Faust, 1994). A directional relationship due to the difference in the roles executed between the actors of the row and the column. The part of the text represented by the actor on the row is the one that cites the part of the text associated with the actor in the column. It is a valued relationship because the cells indicate the number of citations (non-binary) made by the section of text described in the row for the section of text described in the column.

For the software of the AnaCoTEx approach to be able to generate the text matrix of each article, it was necessary to encode each article beforehand. This involved reading the article to identify and encode, in the form of cross-reference, all the interrelations between the different sections and subsections in the articles. A cross-reference is a reference within a text to another part of the text (Collins, 2013). The way a cross-reference is indicated in the text for recognition and analysis by the AnaCoTEx approach software is described at http://www.anacotex.com/en/operacionalizacao.php. In this study, 108 text matrices were generated, one for every article analyzed. An example of one of the generated text matrices is shown in Table 1.

Analyzing the 108 articles involved in-depth reading of the texts to identify the text elements that characterized citations of other internal sections of the articles, to which the cross-reference indicators should be introduced. The forms used by the authors to cite internal texts of the article are highly diverse. They include: a) direct citations of other sections or subsections, e.g., "We describe six cases in detail to illustrate our model in the Results section." (Powell & Baker, 2014; p. 1410); b) citing non-textual elements (tables and figures) from other sections, e.g., "... investigated necessary evils in four occupational settings (see Table A): managers and ..." (Margolis & Molinsky, 2008; p. 850); and c) using terms with identifiers that function as labels to be cited in other sections, as occurs, for instance, with the hypotheses. We present below excerpts from

the articles that illustrate the use of four terms (hypothesis, model, proposition and vignette) that were given an identifying number as a label and were cited in subsequent sections to the section that defined them:

- "... instead will be focused on using the support to protect against further resource loss (see Hypothesis 2)." (Parker, Johnson, Collins, & Nguyen, 2013; p. 874);
- "Deviation from the dominant logic was gradual and moderate (see Model 3), took place at the organizational level ..." (Durand & Jourdan, 2012; p. 1310);
- "... standard deviations above Miller's mean, confirming Proposition 5." (Greenwood, Hinings & Brown, 1990, p.748);
- "As noted in Vignette 1, early in its life, the ..." (Amabile et al., 2001; p. 427).

All these situations received the insertion of cross-references. The process involved five analysts. All the articles were read by two researchers, the first carefully read and encoded the article and the other conducted a new reading and validated the cross-references included in the text of the article.

3.3. Reading of the encoded article and preparation of the text matrix

With the cross-references included in the text of the article, in accordance with the standards described in the previous section, the following step involved reading the encoded text to construct the text matrix. To streamline this activity, the Encoder Software available in the AnaCoTEx approach was used. Every cross-reference identified in the text is marked on the text matrix, with one unit added in the cell associated with the structure of the text with the cross-reference (invoker) and the referenced structure in the brackets of the cross-reference (invoked). The logic for defining the coordinates for the row and column of the cell in the text matrix is that the structure of the text of the invoker is considered as the position of the row of the matrix, and the structure of the invoked text is considered the position in the column of the matrix. For the non-textual elements of the document (figures and tables), the Encoder Software considers them as a paragraph of text, meaning that for this element, no actor is created in the text matrix (it is just another paragraph), linking it as a synonym (nickname) of the section, subsection or other structural level in which the non-textual element is included. Following the reading of the article, when a citation of a previously registered non-textual element is found, the software associates it with the section of text that refers to the structure to which the element is linked as a synonym. If, when reading a paragraph, the encoder software finds a citation of non-textual elements, in the text matrix it is associated with the structure of text in which the paragraph is included, with the nickname/synonym of the structure of text that contains and presents the table or figure in question. The same logic was applied to the terms that served as labels, such as hypothesis, model, proposition and vignette, which are used to indicate a portion of specific text.

After identifying the actors and total values of each cell in the text matrix, the Encoder Software calculates the sums of each of the rows and columns. The total value for each row is referred to as the outdegree centrality of each actor (Wasserman & Faust, 1994), i.e., the interactions that are initiated (invocations made) by the structure of text described in the row for all the other structures of text described in the columns. The total values of each column indicate the indegree centrality of each actor (Wasserman & Faust, 1994), i.e., the interactions received from the other structures of the extensive text. The complete text matrix is generated and presented on the screen at the end of the Encoder Software execution. Only one discrepancy was observed between the logic employed by the AnaCoTEx and the encoded structure in the 108 articles. In Webster and Hackley (1997, p. 1291), in the text of the third section (Method) a distant table, belonging to the fourth section (Result) is mentioned: "Table A (in 'Result') gives internal consistency reliabilities (Cronbach alphas) for all scales". Considering that the location of the first mention of the non-textual element refers to where it is placed in the article, the AnaCoTEx allocated Table A to the Method section and the later citations of Table A as citations of the Method section. Therefore, the text matrix generated for this article had to be altered concerning the citations for Table A.

3.4. Consolidation of the text matrix by types of research

After the text matrices for each of the 108 articles in the sample were created, they were divided into three groups in accordance with the type of research: quantitative, qualitative and qualitative-quantitative. 91 of the 108 articles (84.3%) had five sections in common: Introduction, Literature Review, Method, Results/Findings, Analyses and Discussions. This situation is consistent with the perspective of Sun and Linton (2014, p.571): "Empirical social science journal articles normally consist of six parts: (1) Introduction, (2) Literature review, (3) Methodology, (4) Result, (5) Discussion, and (6) Conclusion". Two alterations were made to this structure to define the standard structure to be used to consolidate the three groups of text matrix: a) the Conclusion sections found in 33 articles (30.5%) of the sample were excluded, as they had no association with the cross-references (the Conclusion sections did not cite and were not cited by other sections of the articles); b) the Appendix section, found in 47 of the 108 articles (43.5%), were added, as many of these articles had one or more appendices, resulting in many cross-references. Thus, all the articles had their cross-references distributed and analyzed in six sections: Introduction, Literature Review, Method, Results/Findings, Analysis and Discussion, and Appendix.

Articles with additional sections or titles different from the standard six-section structure had the content of these sections analyzed and allocated by relevance. In some situations, the distribution was by the assignment of subsections of the same section for different parts of the standard structure. This was the case, for example, for articles that included experiments or

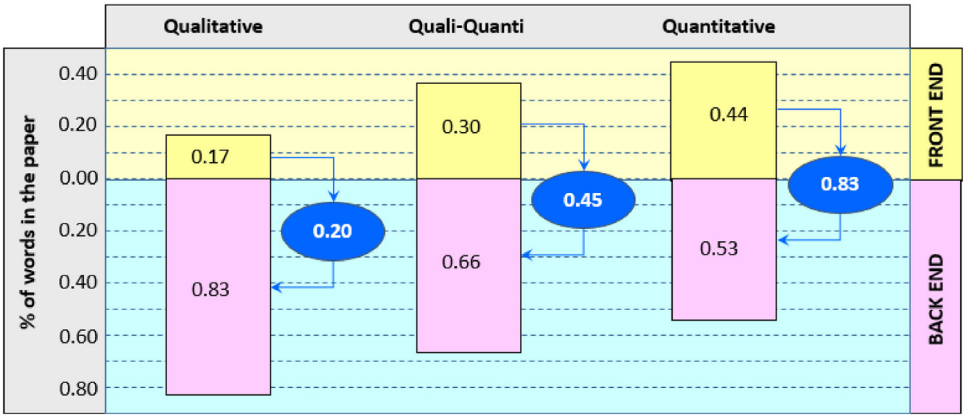

Fig. 1. Size of front end and back end of the articles according to type of research.

quasi-experiments, involving two or more studies, with specific subsections of method and analysis. After the alignment of the text matrix through the standard structure, these were consolidated by type of research, constituting three 36-cell text matrices (6×6 square matrix). In the cells, the total number of articles and the total number of citations were registered. Concomitantly with the consolidation of the 108 text matrices, in accordance with the standard structure, the words were also counted for each of the sections, defining the percentage of each chapter in relation to the whole. For this count, the lists of references of the articles were excluded, as were the figures and tables.

It should be highlighted that the hypothesis formulation process involved both deductive and inductive logic. The first three central hypotheses, which have already been presented, were deduced from editorials focusing on the structuring of scientific articles. Another five hypotheses, introduced in the following section, were induced from the initial set of data collected for the analysis of the first three hypotheses, characterized by 108 text matrices, encompassing 61,042 cells (sum of the cells present in the 108 text matrices which are square matrices, with variable size, depending on the amount of sections, subsections and other subsequent levels of the article, considered as network actors) employed to record the 1532 cross-references identified in the articles. From this database, the concept of Exploratory Data Analysis was applied for the development of new hypotheses (Oquendo et al., 2012).

3.5. Analysis techniques

In this work, whenever possible, the analysis was conducted in accordance with the idea of estimation statistics. Thus, the intention was to avoid the problems pointed out by Cumming (2014). In the cases analyzed in the present study, a formula for a specific confidence interval was adopted for the mean when σ is unknown.

4. Results

4.1. Tabulations and analyses associated with the size of the article sections

To test *H1,* all 108 articles from the sample were considered. To calculate the number of words in the Front End of each article, the words of all the sections that preceded the Method section were added. For the Back End all the words from the method section to the last section were added, without considering the words in the Appendices and References. Fig. 1 shows the arithmetic mean of all the words, described in percentages, calculated for the Front End and Back End of the articles for the three types of research. It also shows the ratio between the average percentage of the total word count for the Front End and the average percentage of words for the Back End. The most equal ratio between the volume of Front End and Back end words (closest to one) was obtained for the quantitative articles, with 0.83. The most unequal ratio (closest to zero) was for the qualitative articles, with 0.20. The ratio with the most intermediate position (closest to the middle) was for the qualitative-quantitative articles, with 0.45.

Hypothesis H1 is validated through the confidence interval (CI) test of the ratios between the volume of Front End and Back End words (*FrontEnd*). Considering that the *FrontEnd* variable assumes values in the interval of 0–1, the following classification criterion was adopted. *FrontEnd* values lower than 0.30 would be for qualitative studies, and over 0.70 would be for quantitative studies, while other values would be for qualitative-quantitative studies (central region). Fig. 2 shows the 95% Confidence Interval for the mean of the ratio between the volume of words of the front end and the back end; this ration can be considered as the discriminatory characteristic of the type of research. There is evidence that on average the ratio between the volume of words of the Front End and Back End (*FrontEnd*) is concentrated in the far left of the interval 0 and 1 for qualitative studies, on the far right for quantitative and in the center for qualitative-quantitative studies, thereby confirming Hypothesis H1.

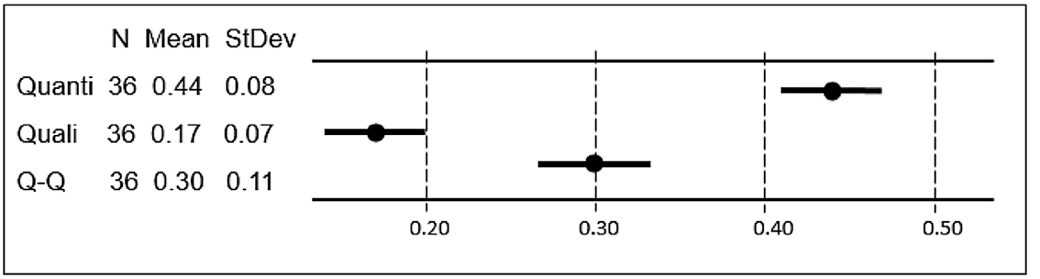

Fig. 2. 95% confidence interval for the mean of the ratio between the volume of words of the front end and the back end.
Legend: Quanti: quantitative studies; Quali: qualitative studies; Q-Q: qualitative-quantitative studies

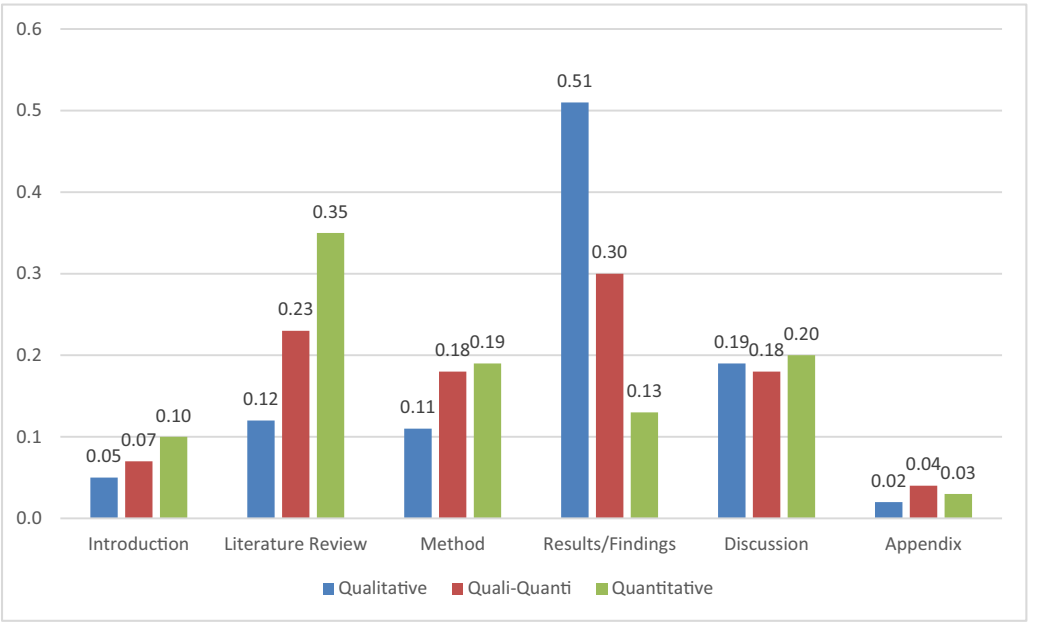

Fig. 3. Size of standard sections of the article according to type of research.

For a comparative analysis between the Front End and Back End of the articles, the size (word count) of each section of the article was calculated, which were associated and consolidated for each of the standard sections of this study. Fig. 3 compares the size of the standard sections according to type of research. An inversely proportional relationship can be seen between the sizes of the standard Results/Findings and Literature Review sections, for both qualitative and quantitative types of research, albeit in reverse order. In qualitative studies, the Results/Findings section is the largest, while in quantitative studies the Literature Review is the largest. In qualitative-quantitative the size of both sections is mostly equal. From these initial tabulations and analysis for the testing of *H1* another relationship was identified with the possibility of characterizing the type of research, which was defined as the fourth hypothesis:

H4–The ratio between the volume of words in the Literature Review section and the Results/Findings section is a discriminatory characteristic of the type of research, whether qualitative, quantitative of qualitative-quantitative, with a higher ratio for quantitative, a lower one for qualitative and, for qualitative-quantitative, a result that lies between the two other types.

Hypothesis H4 is validated by the test comparing the distribution of the ratio between the volume of words in the Literature Review and the Results/Findings section (referred to here as *LiterRes*) between the three types of research (qualitative, qualitative-quantitative and quantitative). See Fig. 4. This figure shows that the ratio between the volume of words in the Literature Review section and the Results/Findings section is a discriminatory characteristic of the type of research.

The greatest difference is between the Quantitative and Qualitative types of research, followed by the Quantitative and Qualitative-quantitative pair. Thus, it is concluded that the ratio between the volume of words in the Literature Review and Results/Findings sections is greater for Quantitative studies, lower for Qualitative and, for qualitative-quantitative, it lies in between the two other types. Furthermore, as the averages of the orders for the types of research are significantly different, it is concluded that the *LiterRes* variable is a potential discriminant factor of the type of research, thereby confirming Hypothesis H4.

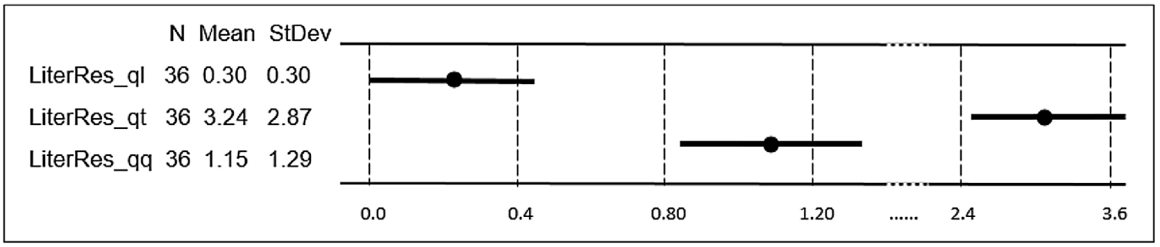

Fig. 4. 95% confidence interval for the mean of the ratio between the volume of words in the literature review section and the results/findings section. Legend: LiterRes: ratio between the volume of words in the Literature Review and the Results/Findings section; ql: qualitative studies; qt: quantitative studies; qq: qualitative-quantitative studies

Table 2
Consolidated text matrix for the 36 qualitative articles.

QUALI	Introduction	Literature Review	Method	Result/ Finding	Discussion	Appendix	OutDegree
Introduction		6 (6 articles)	8 (7 articles)	7 (7 articles)	10 (7 articles)		31
Literature Review		17 (6 articles)	2 (2 articles)			1 (1 article)	20
Method		1 (1 article)	6 (6 articles)	19 (9 articles)	1 (1 article)	9 (6 articles)	36
Result/ Finding	3 (1 article)	7 (3 articles)	40 (6 articles)	**142 (21 articles)**	6 (3 articles)	5 (3 articles)	**203**
Discussion		2 (2 articles)	2 (2 articles)	17 (6 articles)	24 (13 articles)		45
Appendix					.		0
InDegree	3 .	33	58	**185**	41	15	**335**

4.2. Tabulations and analyses of cross-references between sections

The analyses of Hypotheses *H2* and *H3* require the creation of some specific indicators produced by the network analysis technique, such as the indegree and the outdegree. For this purpose, the analysis of these hypotheses are aided by the text matrices generated for each of the three types of research: qualitative, described in Table 2; quantitative, described in Table 3; and qualitative-quantitative, described in Table 4.

H2 mentions the outdegree indicator of the Discussion section for qualitative studies. Thus, in the corresponding text matrix (Table 2), there are 45 cross-references from a total of 335 identified in 36 articles. This same indicator for the other types of research was: 73 citations out of a total of 696, in the 36 quantitative articles (Table 3); 39 citations out of a total of 501, found in the 36 qualitative-quantitative articles (Table 4).

To test Hypothesis H2, a dichotomous variable (*OutDisc*) was defined, which will assume a value of one when in a certain article the number of cross-references in the Discussion section to the others (outdegree) is higher than the number of cross-references in the other sections. The *OutDisc* variable will be a discriminant factor of the type of research if its mean (proportion of occurrences) shows a significant difference between the types of research (groups of research). To compare the proportions between the three types of research the Chi-squared test is used, with a null hypothesis H_0: $p_1 = p_2 = p_3$ and an alternative hypothesis H_1: not all p_i are equal (with i = 1, 2, 3), where p is the proportion of the *OutDisc* variable present in the i-th type of research (Qualitative, Qualitative-quantitative and Quantitative). From a 2×3 contingency table (with two degrees of freedom), the test statistic is given by $\chi^2_{cal} = 3.938$ and p-value = 0.14, so that the null hypothesis H_0 is not rejected. Therefore, at the 0.05 level of significance, it is concluded that there is no evidence that the number of cross-references of the Discussion section to the others is a discriminant factor of the type of research, which does not validate Hypothesis H2.

In the text matrix produced to test *H2* (Table 2), the cell composed of the intersection between Results/Findings and Results/Findings is highlighted by its high values, both for cross-references (142 of a total of 335, 42.4%) and the number of

Table 3
Consolidated text matrix for the 36 quantitative articles.

QUANTI	Introduction	Literature Review	Method	Result/ Finding	Discussion	Appendix	OutDegree
Introduction	2 (2 articles)	17 (8 articles)	9 (7 articles)	5 (4 articles)	5 (3 articles)		38
Literature Review	2 (2 articles)	49 (22 articles)	1 (1 article)				52
Method	1 (1 article)	48 (13 articles)	28 (7 articles)	1 (1 article)	1 (1 article)	39 (15 articles)	118
Result/ Finding	15 (2 articles)	263 (32 articles)	33 (5 articles)	55 (12 articles)	4 (4 articles)	9 (5 articles)	379
Discussion	7 (2 articles)	34 (10 articles)	11 (4 articles)	17 (10 articles)	3 (2 articles)	1 (1 article)	73
Appendix		20 (2 articles)	1 (1 article)	13 (2 articles)	1 (1 article)	1 (1 article)	36
InDegree	27 .	**431**	83	91	14	50	**696**

Table 4
Consolidated text matrix for the 36 qualitative-quantitative articles.

QUALI-QUANTI	Introduction	Literature Review	Method	Result/ Finding	Discussion	Appendix	OutDegree
Introduction		8 (6 articles)	7 (5 articles)	5 (5 articles)	4 (4 articles)	1 (1 article)	25
Literature Review		18 (7 articles)			1 (1 article)	1 (1 article)	20
Method		25 (9 articles)	21 (8 articles)	22 (8 articles)	1 (1 article)	42 (16 articles)	111
Result/ Finding	11 (1 article)	155 (18 articles)	64 (12 articles)	55 (18 articles)	1 (1 article)	16 (7 articles)	302
Discussion		5 (3 articles)	1 (1 article)	29 (8 articles)	3 (2 articles)	1 (1 article)	39
Appendix				1 (1 article)		3 (3 articles)	4
InDegree	11 .	211	93	112	10	64	**501**

articles (21 out of 36, 58.3%). The two analogous cells of the two text matrices registered: 55 citations of a total of 696 (7.9%), produced by 12 of the 36 (33.3%) of the quantitative articles (Table 3); 55 citations of a total of 501 (10.9%), produced by 18 of the 36 (50%) qualitative-quantitative articles (Table 4). From these tabulations and analyses for the testing of *H2*, a more specific aspect of the text matrix was identified, which was capable of discriminating the qualitative type of research better, and this was defined as the fifth hypothesis:

H5–The Results/Findings section of qualitative articles is characterized as having the highest volume of cross-references, i.e., many citations between its own subsections, meaning that this section has the highest outdegree and the highest indegree of all the standard sections of the qualitative article.

Table 5
The results/findings section of qualitative articles has the highest outdegree and the highest indegree of all the standard sections of the qualitative article.

	InDegree			OutDegree		
Categories	O_i	E_i	$(O_i - E_i)^2/E_i$	O_i	E_i	$(O_i - E_i)^2/E_i$
Introduction	3	56	49.99	31	56	11.05
Literature Review	33	56	9.34	20	56	23.00
Method	58	56	0.08	36	56	7.05
Results/Findings	185	56	298.82	203	56	387.90
Discussion	41	56	3.94	45	56	2.10
Appendix	15	56	29.86	0	56	55.83
		$\chi^2_{indegree}$	392.04		$\chi^2_{outdegree}$	486.93

Table 6
Quantitative articles are characterized by the lager volume of cross-references between the results/findings and literature review sections.

	InDegree			OutDegree		
Categories	O_i	E_i	$(O_i - E_i)^2/E_i$	O_i	E_i	$(O_i - E_i)^2/E_i$
Introduction	27	116	68.28	38	116	52.45
Literature Review	431	116	855.39	52	116	35.31
Method	83	116	9.39	118	116	0.03
Results/Findings	91	116	5.39	379	116	596.28
Discussion	14	116	89.69	73	116	15.94
Appendix	50	116	37.55	36	116	55.17
		$\chi^2_{indegree}$	1,065.69		$\chi^2_{outdegree}$	755.19

A more operational way of understanding *H5* is to say that in qualitative studies, the cell of the text matrix with the highest number of citations will be the one identified by the intersection between the row and the column of the Results/Findings standard sections.

To test Hypothesis H5 two variables associated with the network analysis are considered: InDegree and OutDegree. These variables were generated for each of the standard sections: Introduction, Literature Review, Method, Results/Findings, Discussion and Appendix. The idea is to use goodness-of-fit test to compare the distribution of frequencies of these variables (InDegree and OutDegree) with a theoretical even distribution. The Chi-squared test was applied (the statistics are shown in Table 5). Columns four and seven of Table 5 show that the Results/Findings category has greater differences between the number of cross-references observed and the number expected.

Returning to the initial hypotheses formulated from the texts of the editorials, *H3* mentions the indegree indicator of the Literature Review section of the text matrix for quantitative studies. In Table 3, the value of this indicator is shown to be 431 out of 696 cross-references (61.9%). This same indicator for the other types of research had values of 33 cross-references out of 335 (9.85%), for the 36 qualitative articles (Table 2), and 211 out of 501 (42.1%), for the 36 qualitative-quantitative articles (Table 4).

To test Hypothesis H3, a dichotomous variable (*InLiter*) was defined, which will assume a value of one (with probability p) when in a given article the number of cross-references from other sections to the Literature Review section (indegree) is higher than the number of cross-references to the other sections. The *InLiter* variable will be a discriminating factor of the type of research if its mean (proportion of occurrences) shows significant differences between the types of research (groups of research). To compare the proportions between the three types of research, the Chi-squared test is used. Thus, at the 0.05 level of significance, it is concluded that there is evidence that the number of cross-references from the other sections to the Literature Review section is a discriminating factor of the type of research. It should be highlighted that the rejection of the null hypothesis H_0 only permits the conclusion that there is no equality of proportions between the types of research. Therefore, it is not possible to affirm for which pair of groups (types of research) there is difference in the proportions.

From the tabulations and analyses for testing *H3*, a more specific aspect of the text matrix was identified that is capable of discriminating quantitative studies better, and was defined as the sixth hypothesis:

H6–Quantitative articles are characterized by the lager volume of cross-references between the Results/Findings and Literature Review sections, i.e., the Results/Findings section intensely invokes the text of the Literature Review.

A more operational way of understanding *H6* is to say that in quantitative studies, the text matrix cell with the highest number of cross-references will be identified by the intersection of the Results/Findings row and the Literature Review column.

To test Hypothesis H6, two variables associated with network analysis are considered: InDegree and OutDegree. These variables were generated for each of the standard sections: Introduction, Literature Review, Method, Results/Findings, Discussion and Appendix. The idea is to use goodness-of-fit test to compare the distribution of frequencies of these variables (InDegree and OutDegree) with a theoretically even distribution (the proportion of number of cross-references is equal in all the categories. To this end, the Chi-squared test was applied (the statistics of which are shown in Table 6). For each variable, it can be concluded that at least one category has a significantly different number of cross-references than the expected value E = 116.

Table 7
The presence of the term findings as a label of a section of an article is associated with qualitative and qualitative-quantitative studies.

	Quantitative				Qualitative				Qualitative-Quantitative			
	Findings	Result	Both	Others	Findings	Result	Both	Others	Findings	Result	Both	Others
1 st Level – Section	**0**	27	0	6	13	7	0	15	6	15	2	5
2nd Level – Subsection	**0**	0	0	3	0	1	0	0	0	6	1	1
TOTAL	**0**	27	0	9	13	8	0	15	6	21	3	6
	27			9	21			15	30			6

Columns four and seven of Table 6 show that the Literature Review category (in the InDegree variable) and Results/Findings (in the OutDegree variable) have greater differences between the number of cross-references observed and the number expected (E = 116).

The generation of the three text matrices, for analyzing the hypotheses, allowed us to observe further nuances with discriminatory potential for the three types of research. One of the aspects that attracted our attention was the concentration of high values in some regions of the text matrix, with reciprocity between the sections in terms of their cross-references. This led to the application of the Network Analysis core-periphery technique for the three networks described in the three text matrices. A blocked adjacency matrix created by the UCINET (Borgatti, Everett & Freeman, 2002) showed that the sections of the qualitative articles that constitute the central group are Method and Results/Findings. In the case of the quantitative articles, the central group was identified as the Literature review and Results/Findings sections. These central groups were identified by using the calculations of the number of cross-references and the calculation of the number of articles. These results are consistent with the previous findings. In terms of the relationship between the size of the Front end and the Back end, a greater part of the texts of the qualitative articles (81.7%) are concentrated in the Back end sections. Therefore, it was expected that the sections of this part would constitute the central group, as observed in the core-periphery analysis, with the Method and Results/Findings sections. In the case of the quantitative articles, with a better distribution of the text between the Front end (44%) and Back end (53.3%) and a Literature Review section with a very high indegree, the central group was expected to be made up of a section from each part of the article, with the Literature Review representing the Front end. This was indeed the case. Thus, the network analysis core-periphery technique can be considered a tool that aids the identification of the type of research, as there is the text matrix generated for the article.

4.3. Other characteristics identified following the tabulation of the three text matrices

The consolidation of the 108 text matrices into three groups according to the type of research resulted in Tables 2–4, and provided evidence of two other characteristics with a discriminatory potential regarding the type of research, in addition to those initially abstracted from the editorials. These aspects are the labels used in sections and subsections of text and the total number of cross-references in the articles. These two characteristics are described in the following two paragraphs.

4.3.1. Labels of sections and subsections of text

To analyze the cross-references, the titles of the different sections and subsections of the article were encoded and used to denominate each of the actors of the text matrices. Following the identification of the titles in the text matrices and their later tabulation, one aspect stood out: the recurrence of the label "Result" and the label "Findings", identified in 78 articles in the sample (72.2%). Table 7 shows the distribution of these 78 labels for the three types of research. The Results label is indistinctly present in all three types of research, whereas the Findings label is used only in qualitative and qualitative-quantitative studies. The term Results is not representative of any of the three types of research, considering that it is employed indistinctly in all three. As for the term Findings, it is only applied correctly in qualitative or qualitative-quantitative studies. As it is not used improperly in quantitative studies, it can assume that there is evidence that the detection of Findings as a label on one of the sections of the article precludes the possibility of the article being classified as quantitative. This led us to draft the seventh hypothesis:

H7–The presence of the term Findings as a label of a section of an article is associated with qualitative and qualitative-quantitative studies, precluding it from being a quantitative study.

For the purposes of testing hypothesis H7, two categorical variables were considered: Presence of the Findings label, with Yes and No categories, and the Type of Research variable, with Quantitative and Non-quantitative categories (Qualitative or Qualitative-quantitative). The idea is to use the Chi-squared test for independence with the aim of evaluating the degree of association between these variables, to enable us to verify whether the term Findings is related to the type of research, i.e., if these variables are dependent.

4.3.2. Number of cross-references in the article

The number of cross-references identified in the qualitative, qualitative-quantitative and quantitative articles, respectively, was 335, 501 and 696, as described in Tables 2–4. This led us to draft the eighth hypothesis:

H8–The number of cross-references in quantitative articles is higher than the number of cross-references in qualitative articles.

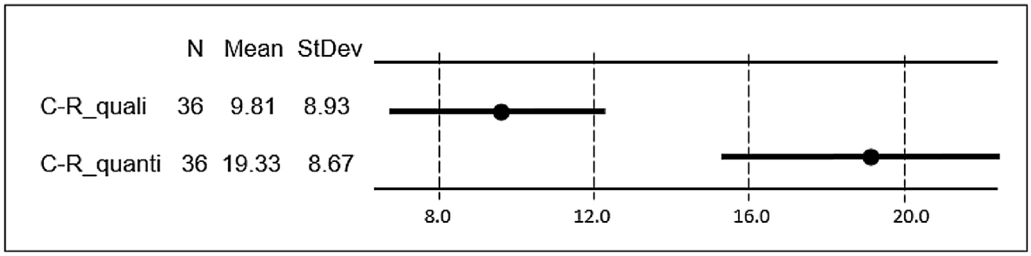

Fig. 5. 95% confidence interval for mean of the number of cross-references in quantitative articles and qualitative articles.
Legend: C-R: number of cross-references; quali: qualitative studies; quanti: quantitative studies

Hypothesis H8 is validated from a test that compares the distribution of the number of cross-references (*CrossRef*), among the three research groups (qualitative, qualitative-quantitative and quantitative). The results are presented in Fig. 5, which shows that the number of cross-references in quantitative articles is higher than the number of cross-references in qualitative articles. Therefore, it is concluded that the number of cross-references is higher for Quantitative studies than for Qualitative, thus confirming Hypothesis H8.

5. Analyses and discussions

In this article, we explored the structural aspects of scientific articles capable of characterizing them according to the type of research (qualitative, quantitative or qualitative-quantitative). There are three kinds of analyzed variables: a) those related to the size of the sections of the article (word count); b) those related to the cross-references between sections, subsections and other levels of the structure of the article; and c) those related to the label of sections. In addition to identifying the structural characteristics of scientific articles according to the type of research, this study proposes to discuss the utility of these characteristics for the purposes of classification, analysis and development of scientific articles. We will begin the discussion in terms of classification activity.

Among the recommendations of Langley (1999) on the dilemma of dealing with the accuracy, generality and simplicity dimensions in the production of scientific knowledge, it can be highlighted that "simple theories with good explanatory power may actually be preferred to complex ones that explain a little more" (Langley, 1999; p. 695).

The eight variables are associated with eight hypotheses, three of which were directly derived from the editorials (H1, H2 and H3), while the others were derived from the data generated to test the first three. Seven of these hypotheses were corroborated by the tests, which enabled us to infer twelve rules associated with the structure of a scientific article according to its type of research, with six associated with quantitative studies (R1-R6), four with qualitative (R7-R10) and two with qualitative-quantitative (R11-R12), as shown in Table 8. These rules can act as parameters of verification for those who develop and evaluate scientific articles.

Six of the twelve rules are associated with the ratio between the number of words (R1, R2, R7, R8, R11 and R12), either between parts of the text (H1) or between sections (H4). H1 compares the volumes of the Front end and Back end. H4 compares the Literature Review section with the Results/Findings section. The empirical tests conducted in this study for H1 validate the comments and opinions of editors Bansal and Corley (2012), as expressed in their editorial. The logic of H1 presents a comparison of larger portions of text (parts) which had their words added, but are more easily identified due to the fact that front end and back end parts are adjacent, i.e., one part ends and the other begins at the same point. In H4, there is a smaller volume of words for only two of the standard sections, the Literature Review and Results/Findings, but with a more difficult identification, considering that there is no common division point, as occurs in H1, but four points, delimiting two sections that are normally not adjacent.

To choose the rules associated with the number of words, obtained from H1 and H4, the idea is to apply the procedure that is easiest and simplest to the context of each article or each collection of articles (repository). The rules associated with H1 (R1, R7 and R11) can be applied to the context of an isolated article with no need for other prior indicators associated with a larger set (collection or repository) as a parameter for comparison. For the rules associated with H4 (R2, R8 and R12), there should already be a prior set of classified articles with a history of previously generated structural indicators. This will enable a relationship to be established for the indicators of each new article to be analyzed and classified by comparing the available indicators for each type of research. This same principle of historical indicators consolidated for a set of articles in a journal or repository to be used for the analysis parameter of the scientific article, applies to rules R6 and R10. Therefore, we have five rules whose application should be associated with the context of the environment (mean indicators of the journal, collection or repository) and a further seven that can be applied in isolation, article by article, or to groups of articles. This information is described in the last column of Table 8.

An analysis of the value of the ratio obtained from the number of words in parts of the text is an important aspect for analyzing the quality of a scientific article. The importance of this topic is highlighted in the introduction of the editorial by Bansal and Corley (2012, p.509), in which they address the volumetric relationships between parts of a qualitative article: "This editorial concludes a seven-part series, 'Publishing in AMJ', in which the editors give suggestions and advice

Table 8
Rules associated with the structure of the business management articles.

Id Rule	Type of Research	Hypothesis	Distinctive Aspect	Description of Rule	Application of Rule
R1	Quant	H1	Number of words	The ratio between the number of words of the Front end and Back end (*FrontEnd*) for Quantitative studies is on average located at the far right of the open interval (0, 1), higher than 0.7	Isolated and Contextual
R2	Quant	H4	Number of words	The ratio between the volume of words in the Literature Review and the Results/Findings (*LiterRes*) is higher for Quantitative research than the values found for the Qualitative studies and the Qualitative-quantitative studies	Contextual
R3	Quant	H3	Cross-References	The Literature Review section has the highest indegree (*InLiter*) of the six standard sections of the Quantitative research	Isolated and Contextual
R4	Quant	H6	Cross-References	Of the 36 cells of the text matrix, the one constituted by the intersection between the standard section of Results/Findings (row) and the Literature Review (column) (*ResxLiter*) has the highest value of cross-references for Quantitative research	Isolated and Contextual
R5	Quant	H7	Label	The presence of the term Findings as a label (*TitFind*) of a section or subsection of an article precludes it from being Quantitative research	Isolated and Contextual
R6	Quant	H8	Cross-References	Quantitative studies have more cross-references (*CrossRef*) than Qualitative studies	Contextual
R7	Quali	H1	Number of words	The ratio between the volume of words in the Front end and Back end (*FrontEnd*) for Qualitative studies is on average located at the far left of the open interval (0, 1), lower than 0.3	Isolated and Contextual
R8	Quali	H4	Number of words	The ratio between the volume of words in the Literature Review and the Results/Findings (*LiterRes*) section is lower for Qualitative studies in comparison with the values found for the Quantitative studies and the Qualitative-quantitative studies	Contextual
R9	Quali	H5	Cross-References	Of the six standard sections, the section on Results/Findings has the highest value for the outdegree indicator and also the highest value for the indegree (*ResxRes*) indicator for Qualitative research	Isolated and Contextual
R10	Quali	H8	Cross-References	Qualitative research has less cross-references (*CrossRef*) than quantitative studies	Contextual
R11	Quali- Quant	H1	Number of words	The ratio between the volume of words in the Front end and Back end (*FrontEnd*) for Qualitative-quantitative research is on average located in the interval between 0.3 and 0.7	Isolated and Contextual
R12	Quali- Quant	H4	Number of words	The ratio between the volume of words in the Literature Review and the Results/Findings (*LiterRes*) section for Qualitative-quantitative studies is higher than the values found for the Qualitative studies and lower than the values found for the Quantitative studies	Contextual

for improving the quality of submissions to the Journal". A ratio value other than expected, in accordance with the type of research, can be obtained because of problems involving: a) the simplification of one of the parts of the text, which is shorter (less detailed) than it should be; or b) an excess of words in one of the parts, resulting in fastidiousness and prolixity. Thus, the ratio between the parts of the text appears to be an important characteristic when it comes to gauging the quality of a scientific article.

The word count has been the resource used as a structural aspect for analyzing how well scientific articles conform to the standard for their particular type of research. This study innovates by incorporating network analysis indicators, such as the indegree and outdegree, as a resource for gauging the structural conformity of the article to its type of research. Of the twelve rules identified, five are associated with indicators generated from the analysis of cross-references between sections and subsections of the article. Of these five rules, three are associated with quantitative studies (R3, R4 and R6) and two with qualitative (R9 and R10). Obviously, the indicators associated with cross-references are more difficult to gauge in relation

Table 9
Variables and indicators for the structural analysis of scientific articles.

Indicator	Formula	Variables	Reference Values for Management	
TR1: Type of research1	TR1 = Vwfe/Vwbe	(Vwfe) volume of words of the front end (Vwbe) volume of words of the back end	**TR_1** $TR_1 < 0.30$ $0.30 \leq TR_1 < 0.70$ $TR_1 \geq 0.70$	**Type of research** Qualitative research Qualitative-quantitative research Quantitatitive research
TR2: Type of research2	TR2 = Vwlr/Vwrf	(Vwlr) volume of words in the Literature Review (Vwrf) volume of words in the Results/Findings	**TR_2** $TR_2 < 0.40$ $0.40 \leq TR_2 < 1.50$ $TR_2 \geq 1.50$	**Type of research** Qualitative research Qualitative-quantitative research Quantitatitive research
$Cqlr_1$: confirmation of qualitative research$_1$	$Cqlr_1$ = OutDisc	(OutDisc) outdegree of the Discussion	Cqlr1 > 4.83[a]	
$Cqlr_2$: confirmation of qualitative research$_2$	$Cqlr_2$ = Inrf + Outrf	(Inrf) indegree of the Results/Findings (Outrf) outdegree of the Results/Findings	Cqlr2 > 9.23[a]	
$Cqtr_1$: confirmation of quantitative research$_1$	$Cqtr_1$ = InLiter	(InLiter) indegree of the Literature Review	Cqtr1 > 9.80[a]	
$Cqtr_2$: confirmation of quantitative research$_2$	$Cqtr_2$ = RfiLiter	(RfiLiter) Results/Findings invokes the text of the Literature Review	Cqtr2 > 4.11[a]	
C ∼ qtr: confirmation of non-quantitative research = confirmation of qualitative research or qualitative-quantitative research	C ∼ qtr = FasLabel	(FasLabel) presence of the term Findings as a label of a section	C ∼ qtr > 0	

[a] The reference values are equal to 0.5 of the average values.

to the other indicators associated with the word count, as it is necessary to encode each point of citation and reference to internal texts to generate the text matrix. However, it should be highlighted that over 80% of the cross-references of the articles in the sample were the result of mentioning non-textual elements (tables and figures) and labels like those used to identify hypotheses and sections of text. These cross-references are simple to identify automatically using software for reading and generating the text matrix, as is the case of the AnaCoTEx used in this study. It is important to note that the introduction of variables associated with cross-references increases the accuracy of the analysis.

The use of network analysis to analyze the internal structure of texts can trigger other perceptions and knowledge associated with the quality of scientific articles. A point in question, in the present study, is the tabulation of the data necessary to assemble and operate the text matrix for the analysis of cross-references, which led to the perception of the Findings label as a discriminatory characteristic of the qualitative research (R5). It was observed that the great tradition of the positivist study led to the Results label being used indiscriminately through force of habit, irrespective of the type of research. Meanwhile, the term Findings, used to designate scientific knowledge associated with more recent research paradigms, has been used more sparingly and much more assertively in relation to the type of research.

To conclude, let us return to the aim of this study, which is to identify variables and indicators for the development of rules focusing on the structural analysis of scientific articles. In the fourth column of Table 8, the proposed variables for the structural analysis of texts are shown: cross-references, number of words, and label. The variables abstracted from the theoretical framework are also added: parts of the text (front end, back end), type of research (qualitative, quantitative and qualitative-quantitative) and standard sections of the text document (Introduction, Theory/Literature-review, Method, Results/Findings, Discussion and Appendix). To facilitate the understanding and application of the indicators, in Table 9 we present a description of all the proposed variables and indicators.

Of the variables identified for generating the indicators, we consider cross-references as the most significant contribution, not only due to the results obtained in the discriminant analysis technique, but by the unprecedented nature of this resource when applied to the structural analysis of texts. The procedures to generate the network analysis indicators (indegree and outdegree), and the development of the text matrices from the cross-references are described in the first four subsections of the Section 3.

The analysis of cross-references is promising in terms of analyzing the quality of scientific articles. We view it as a source of information that is just as important to science as external references. The external references enabled a series of studies, especially in the field of bibliometrics, associated with the quality of scientific research and of those involved in it. Scientific journals and repositories of scientific articles have algorithms that, at the time of uploading articles submitted for analysis, read, encode and tabulate the list of external references of the article to provide analytical information to editors and other interested parties. Likewise, algorithms for analyzing cross-references can provide useful information to authors and editors

on the quality of the article, identifying, for instance, nonconformities between the type of research and the indicators obtained through a structural analysis of the article.

In the same way that the external reference indicators have their application contextualized for each field of science due to the variation of the number of citations per article and the obsolescence rate in each field (Huth, 2001), the indicators associated with the internal references should also be contextualized in accordance with the field of interest. Thus, the measurements attributed to the different indicators (or reference values) should be calculated for each repository or set of journals devoted to a field of science, for which the development of the rules is intended to support the writing, classification and analysis of scientific articles.

To continue this line of research, we suggest using the text matrix to analyze characteristics of scientific articles associated with other taxonomies of scientific research, such as: a) different type of alternative knowledge claims, including Postpositive, Socially constructed, Advocacy-participatory and Pragmatic (Creswell, 2003); or b) the different types of research strategies associated with a specific research paradigm, such as Phenomenology, Ethnography and Grounded Theory, associated with the socially constructed paradigm (Creswell, 2003).

Proposition of Indicators for the Structural Exegesis of Scientific Articles

Authors' contribution

José Osvaldo De Sordi: Conceived and designed the analysis, collected the data, wrote the paper.

Wanderlei Lima de Paulo: Conceived and designed the analysis, Contributed data or analysis tools; He tested the eight hypotheses. Perform analysis.

Manuel Antonio Meireles: He tested the eight hypotheses. Perform analysis, wrote the paper.

Marcia Carvalho de Azevedo: Collected the data, wrote the paper, Contributed data or analysis tools; She was responsible for executing AnaCoTEx software for each of sample articles and generate the 108 Text Matrix.

Luis Hernan Contreras Pinochet: Collected the data; He made the identification and encoding of cross-references between chapters and sub-chapters of each article under study.

References

Amabile, T. M., Patterson, C., Mueller, J., Wojcik, T., Odomirok, P. W., Marsh, M., et al. (2001). Academic-practitioner collaboration in management research: A case of cross-profession collaboration. *Academy of Management Journal*, *44*(2), 418–431.

Bansal, P., & Corley, K. (2012). From the editors –Publishing in AMJ – part 7: What's different about qualitative research? *Academy of Management Journal*, *55*(3), 509–513.

Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2002). *Ucinet for windows: Software for social network analysis*. Harvard, MA: Analytic Technologies.

Cargill, M., & O'Connor, P. (2009). *Writing scientific research articles: Strategy and steps*. Chichester, UK: Wiley-Blackwell.

Collins. (2013). *English dictionary*. Retrieved from:. http://www.collinsdictionary.com/dictionary/english/cross-reference

Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches* (2nd Ed.). Thousand Oaks, CA: Sage Publications.

Cumming, G. (2014). The new statistics: wWhy and how. *Psychological Science*, *25*(1), 7–29.

De Sordi, J. O., Meireles, M., & Oliveira, O. L. (2016). The Text Matrix as a tool to increase the cohesion of extensive texts. *Journal of the Association for Information Science and Technology*, *67*(4), 900–914.

Duggan, G. B., & Payne, S. J. (2009). Text skimming: The process and effectiveness of foraging through text under time pressure. *Journal of Experimental Psychology: Applied*, *15*(3), 228–242.

Durand, R., & Jourdan, J. (2012). Jules or Jim: Alternative conformity to minority logics. *Academy of Management Journal*, *55*(6), 1295–1315.

Greenwood, R., Hinings, C. R., & Brown, J. (1990). 'P2-form' strategic management: Corporate practices in professional partnerships. *Academy of Management Journal*, *33*(4), 725–755.

Huth, E. J. (2001). Authors, editors, policy makers, and the impact factor. *Croatian Medical Journal*, *42*(1), 14–17.

Joy, A., & Sherry, J. F. (2003). Speaking of art as embodied imagination: A multisensory approach to understanding aesthetic experience. *Journal of Consumer Research*, *30*(2), 259–282.

Langley, A. (1999). Strategies for theorizing from process data. *Academy of Management Review*, *24*(4), 691–710.

Linton, J. D. (2012). How to get your papers rejected (or not). *Technovation*, *32*(1), 6–8.

Margolis, J. D., & Molinsky, A. (2008). Navigating the bind of necessary evils: Psychological engagement and the production of interpersonally sensitive behavior. *Academy of Management Journal*, *51*(5), 847–872.

Oquendo, M. A., Baca-Garcia, E., Artés-Rodríguez, A., Perez-Cruz, F., Galfalvy, H. C., Blasco-Fontecilla, H., et al. (2012). Machine learning and data mining: Strategies for hypothesis generation. *Molecular Psychiatry*, *17*(10), 956–959.

Parker, S. K., Johnson, A., Collins, C., & Nguyen, H. (2013). Making the most of structural support: Moderating influence of employees' clarity and negative affect. *Academy of Management Journal*, *56*(3), 867–892.

Powell, E. E., & Baker, T. (2014). Itís what you make of it: Founder identity and enacting strategic response to adversity. *Academy of Management Journal*, *57*(5), 1406–1433.

Singh, A., & Singh, S. (2006). What is a good editorial? *Mens Sana Monographs*, *4*(1), 14–17.

Sparrowe, R. T., & Mayer, K. J. (2011). From the editors – publishing in AMJ – part 4: Grounding hypotheses. *Academy of Management Journal*, *54*(6), 1098–1102.

Sun, H., & Linton, J. D. (2014). Structuring papers for success: Making your paper more like a high impact publication than a desk reject. *Technovation*, *34*(10), 571–573.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge, UK: Cambridge University Press.

Webster, J., & Hackley, P. (1997). Teaching effectiveness in technology-mediated distance learning. *Academy of Management Journal*, *40*(6), 1282–1309.

Zhang, Y., & Shaw, J. D. (2012). From the editors –Publishing in AMJ – part 5: Crafting the methods and results. *Academy of Management Journal*, *55*(1), 8–12.