

Publication status: This preprint has not been published elsewhere.

Communicative coherence in management accounting: theory and practice in an automated pipeline

Renato Gomes de Oliveira, Fabrícia Silva da Rosa

<https://doi.org/10.1590/1808-057x20262435.en>

Submitted on: 2026-07-22

Posted on: 2026-07-22 (version 1)

(YYYY-MM-DD)

DOI: 10.1590/1808-057x20262435.en

e-location: e2435

Communicative coherence in management accounting: theory and practice in an automated pipeline

Renato Gomes de Oliveira¹

<https://orcid.org/0009-0001-2366-0808>

Email: renato.go@posgrad.ufsc.br

Fabricia Silva da Rosa¹

<https://orcid.org/0000-0003-4212-1065>

Email: fabricia.rosa@ufsc.br

¹ Universidade Federal de Santa Catarina, Centro Socioeconômico, Departamento de Ciências Contábeis, Florianópolis, SC, Brazil

Received on 08/03/2025 – Desk acceptance on 08/12/2025 – 4th version approved on 02/12/2026

Academic Editor-in-Chief: Andson Braga de Aguiar 

Associate Editor: Cláudio de Araújo Wanderley

This is a preprint of an accepted manuscript. This version has been peer-reviewed, accepted for publication. It is currently undergoing final editing, and will be available under the DOI <https://doi.org/10.1590/1808-057x20262435.en>

ABSTRACT

In light of the relevance, text genre, and signaling theories, this study investigates the extent to which communicative coherence (CC) between the titles and abstracts of management accounting (MA) articles varies by subfield and journal tier. There is a lack of theoretical and methodological consensus on how to measure CC in MA and link it to impact metrics in a comparable way. The evidence is fragmented and heterogeneous in terms of definitions and procedures, which prevents cumulative syntheses. To address this issue, we developed an automated metric and analyzed 13,614 MA articles indexed in Scopus from 1995 to 2025. We applied natural language processing techniques in an automated pipeline. Methodologically, the study constructed a thematic dictionary of MA and developed a replicable protocol for semantic analysis to ensure transparency, reproducibility, and adherence to open science principles. Our analysis revealed that 8.7% of articles exhibited high CC, 62.9% exhibited medium CC, and 28.4% exhibited low CC. Articles from established subfields and those published in Q1–Q2 journals tended to have a higher number of citations associated with high CC. However, the differences between the medium and high levels were not statistically significant. These results suggest that textual clarity is a sign of the field's discursive maturity, establishing CC as an indicator of integrity and readability in academic writing. Our findings contribute to improving editorial and educational practices by demonstrating that consistency between the title↔abstract tends to facilitate understanding of the study's scope and guides screening processes without directly affecting interest, visibility, or citations. This practical contribution strengthens Sustainable Development Goal 4 (SDG 4) (Quality Education) by promoting the clear, accessible, and socially engaged dissemination of scientific knowledge.

Keywords: communicative coherence, natural language processing, scientific communication, SDG 4, management accounting.

1. INTRODUCTION

Traditionally, management accounting (MA) has covered technical topics such as performance evaluation, costs, budgeting, and control systems (Ferreira & Otley, 2009; Merchant, 2010). However, the field has been criticized for its conceptual gaps and inconsistent methodological rigor (Luft & Shields, 2003; Malmi & Granlund, 2011), particularly in comparison to financial accounting (Shields, 2015).

Ibero-American studies have documented the growth of MA research over the past decade (Lunkes et al., 2012; Lunkes, Ripoll-Feliu, & Rosa, 2011). The literature indicates that textual production shapes accounting practices and teaching (Kloppel et al., 2012) supports transparency in knowledge construction (Lacey, 2011), aligning with open science values in Brazil (Russo, 2014).

Recently, the literature has expanded to include themes of innovation, sustainability, and digital transformation in MA (Barreto et al., 2024; Zafar et al., 2024). A recent systematic review documents an increase in the volume of MA articles published between 2003 and 2022, as well as an increased emphasis on theoretical grounding (Beuren & Souza, 2025). This trend aligns with the third mission of universities (Compagnucci & Spigarelli, 2020), which emphasizes engagement with society, methodological transparency, and open access.

Therefore, the objective of this study is to automatically measure communicative coherence (CC) between the title↔abstract of 13,614 MA articles indexed in Scopus from 1995 to April 2025. The study will also test (i) how CC varies by subfield and editorial tier (SCImago Journal Rank [SJR]) and (ii) the association between CC levels and citations in light of the relevance, text genre, and signaling theories. The proposed systematization consolidates scattered evidence into a single framework, revealing comparative patterns of textual organization in the field. It also offers readers a comprehensive reference for understanding the structure of scientific communication in accounting literature.

The contribution is threefold. First, on a theoretical level, the study positions CC as a textual indicator associated with communicative quality in accounting literature. Second, methodologically, it offers a replicable protocol for semantic analysis supported by natural language processing (NLP) and artificial intelligence (AI). Finally, in practice, it provides authors, reviewers, and editors with diagnostic insights to identify and reflect on inconsistencies between title↔abstract. This concern also aligns with Sustainable Development Goal 4 (SDG 4) (Quality Education) because more informative, aligned titles and abstracts enhance the accessibility and pedagogical use of scientific results.

2. THEORETICAL FRAMEWORK AND HYPOTHESES

The CC construct is based on three complementary frameworks: (i) relevance theory (Wilson & Sperber, 2004) – the alignment between the title↔abstract should maximize the cognitive effect with minimal effort; (ii) text genre theory (Swales, 1990) – the title↔abstract pair should mirror the article’s rhetorical moves; and (iii) signaling theory (Spence, 1973) – textual cues in the title↔abstract textual pair function as verifiable signals that reduce uncertainty and anchor interpretation.

Together, these perspectives establish CC as a textual property of title↔abstract alignment (Figure 1).

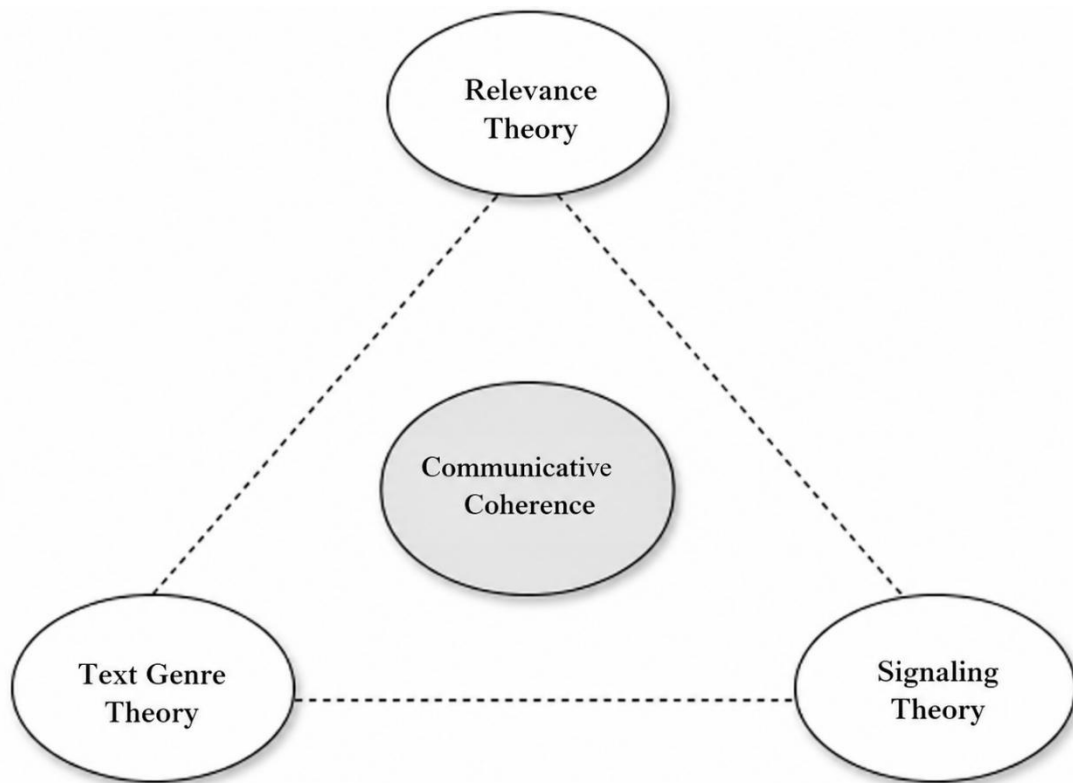


Figure 1 *Theoretical model of communicative coherence*

Source: *Prepared by the authors.*

We do not propose a unified theory. The originality lies in the automated operationalization of the construct. We treat informational expectations as an *ex ante* textual property and verify whether the elements signaled in the title are reflected in the abstract. Section 3.2 details the measurement of CC, and Section 3.3 details the classification criteria.

Within this framework, H₁ relates to relevance (variation in CC by subfield), H₂ relates to text genres (differences in CC by editorial tier [SJR]), and H₃ relates to signaling (association between CC levels and citations).

2.1 MA: Conceptual Foundations

Since the 1990s, MA has expanded beyond its traditional role to include themes such as strategy, innovation, and organizational transformation (Luft & Shields, 2003). However, this diversification has drawn criticism for its functionalist bias (Scapens, 2006) and the difficulty of achieving conceptual innovation (Merchant, 2010). Recent studies have identified more than 290 distinct theories employed in the field (Beuren & Souza, 2025), which signals vitality, but also epistemological fragmentation (Plesner, 2024; Shields, 2015).

The role of budgeting has evolved from a technical tool to a strategic management instrument. Otley (1978) anticipated the behavioral effects of goals, and Shields (1989) and Merchant (2007) emphasized its strategic applications. Similarly, cost management has become integrated into organizational control systems (Malmi & Brown, 2008).

The literature on management control systems (MCS) advanced by articulating control and innovation. Since Ansari's pioneering studies on contingency chains in 1977,

the field has incorporated more dynamic and integrative approaches. Simons (1995) introduced levers of control, while Chapman (1997) emphasized the interaction between formal and informal controls in decision making. Ferreira and Otley (2009) broadened the perspective by presenting a framework connecting control mechanisms, strategy, and the environment. Henri (2006) and Malmi and Brown (2008) emphasized balancing diagnostic and interactive controls. This point was reinforced by Müller-Stewens et al. (2020), who linked this balance to innovation and adaptability.

Performance evaluation has become a cornerstone of MA. Ittner and Larcker (1998) demonstrated how goals influence managerial attitudes, and Otley (1999) integrated financial and non-financial measures. Lillis (2002) discusses the trade-offs between efficiency and flexibility, and Abernethy and Brownell (1999) and Birnberg et al. (2006) demonstrate how cognitive and social factors influence the use of controls.

In recent decades, the scope of MA has broadened. In environmental accounting, the role of socio-environmental information in establishing legitimacy has become more prominent (Deegan, 2002; Gray, 2010). Regarding environmental, social, and governance (ESG) issues, studies have discussed organizational implications (Rodrigue et al., 2013; Zhang et al., 2020). Corporate governance is analyzed through multiple institutional mechanisms (Krishnamoorthy et al., 2004) and ethical implications (Endenich & Trapp, 2020). The debate on innovation incorporates dynamic capabilities (Scarpellini et al., 2020) and performance measurement using big data (Aboagye-Otchere et al., 2021). Digital transformation is reshaping professional competencies (Savić & Pavlović, 2023) and integrating AI into accounting analysis (Abbas, 2025).

The evolution of MA is also influenced by the educational context. The formation of “academic tribes” (Becher & Trowler, 2001) fragments approaches and hinders the construction of integrative paradigms (Lillis, 2002). In Brazil, accreditation processes and curriculum standards have accentuated segmentation (Kloppel et al., 2012; Lunkes, Rosa, Gasparetto, & Balduino, 2011), alongside initiatives for pedagogical innovation (Miller, 2021). Incorporating sustainability and social justice into education reflects openness to critical approaches (Alawattage & Wickramasinghe, 2024). This diversity affects academic communication (Becher & Trowler, 2001), and shapes discursive styles and patterns (DiMaggio & Powell, 2005).

Debates about conceptual clarity and methodological rigor accompany this evolution, as reflected in the article’s paratext. In systematic review protocols, initial screening based on the title and abstract is a standard step for determining eligibility (Ensslin et al., 2013; Page et al., 2021). Vague titles can undermine the visibility and perceived value of an article (Hyland, 2004). When the promise (title) and the delivery (abstract) do not align, the retrieval process in reviews may be less accurate, increasing the risk of false positives and false negatives (Hyland & Tse, 2005). In this context, misalignments between the title↔abstract do not negate the usefulness of the study, but they increase the cost of screening and the risk of early rejection. The title, abstract, and keywords determine the article’s discoverability and visibility as entry points. Their informational alignment guides the reader’s decision to access the full text and potentially impacts the reach of the work (Garcia et al., 2019).

In light of relevance theory (Wilson & Sperber, 2004), these historical-thematic foundations of MA support the view that title↔abstract CC should be understood as a balance between cognitive effect (informational gain) and processing effort (interpretive cost). For analytical purposes, we define established subfields as those with a continuous presence and persistent recurrence in the literature over the period, and maturing subfields as those with more fluid conceptual boundaries or the most recent fronts. As additional bibliographic evidence, we consider terminological stability and centrality in field

syntheses without making normative judgments. Based on this, we postulate the following:

H_{1a}: established subfields of MA are associated with higher CC between the title↔abstract.

H_{1b}: maturing subfields of MA are associated with lower CC between the title↔abstract.

2.2 Scientific Communication and Integrity

The social role of universities is expanding due to open science policies, editorial guidelines, and evaluation criteria. Consequently, accountability in scientific communication is becoming increasingly important. The third mission of universities extends beyond technology transfer to include the creation of public value from knowledge (Compagnucci & Spigarelli, 2020). In this context, CC serves as an observable textual indicator that supports transparency and editorial curation.

The concept of CC is also aligned with the challenges faced by research in MA, as discussed by Merchant (2010). MA research is often pressured to be empirically rigorous and focused on quantitative data and economic models. However, Merchant suggests that the diversity of methods and topics in accounting research is essential for the discipline's advancement. Thus, the relationship between the title and abstract reflects the tension between empirical rigor and the need for a broader approach, which may influence perceptions of relevance in the field of MA.

Evidence regarding paratexts reinforces this interpretation. More assertive titles are associated with more downloads and citations (Jamali & Nikzad, 2011; Sangster, 2015), while structured abstracts promote engagement and understanding of the scope (Hyland & Tse, 2005). Titles and abstracts are complementary genres; the title summarizes the topic, while the abstract presents the argument in a concise manner (Garcia et al., 2019). Decisions to read and cite a paper can be influenced by the rhetorical impact of these paratexts (Burton et al., 2022). In high-impact journals, greater editorial pressure for clarity and standardization reinforces these patterns (Sangster, 2015).

In light of genre theory (Swales, 1990), the title and abstract are rhetorical devices that reflect the study's promise and anchor its interpretation when they are structurally and semantically symmetrical. Genre competence is defined as mastery of the structural conventions of the scientific genre, such as an informative title and an abstract that mirrors the objectives, methods, and findings through recognizable patterns. Genre performance is the concrete realization of these conventions in a specific context, taking into account rhetorical purpose, audience, editorial norms, and authorial style (Devitt, 2015). This distinction allows for a non-punitive interpretation; inconsistencies between the title and abstract are not necessarily "errors" of competence, but rather, may reflect discursive choices and institutional contexts. However, as Merchant (2010) observes, accountability in academic research must also be understood from the perspective of broad communication. This perspective reflects the multiple facets of MA, ranging from traditional practices to new methodologies and interdisciplinary approaches.

In this study, CC indicates paratextual alignment between the promise (title) and delivery (abstract) without evaluating style, originality, or authorial voice. The journal's impact is indicated by the SJR, which classifies journals by field into quartiles: Q1 = top 25% (75th–100th percentile), Q2 = 25th–50th percentile, Q3 = 50th–75th percentile, and Q4 = bottom 25% (0th–25th percentile). For analytical purposes, we grouped Q1–Q2 as

“higher impact” and Q3–Q4 as “lower impact,” assessing title↔abstract CC exclusively between these tiers and positing the following:

H_{2a}: articles published in higher-impact journals (Q1–Q2) exhibit higher CC between the title↔abstract.

H_{2b}: articles published in lower-impact journals (Q3–Q4) exhibit lower CC between the title↔abstract.

2.3 Signaling and Scientific Value

Scientific communication operates under informational asymmetries. In the initial screening process, many engagement-related decisions are made based on paratexts. In this scenario, the title↔abstract are not just labels; they act as signals that shape perception of value and influence access to the full text.

In light of signaling theory (Spence, 1973), credible signals incur costs – such as time, effort, rigor, and editorial risk – that enhance the credibility of the message. Clear textual signals facilitate discovery and use in search engines and indexing systems, thereby increasing the visibility of the work (Colombo, 2021; Venturini et al., 2022). According to Merchant (2010), the hard sciences have an advantage in terms of the clarity and definition of academic signals, making communication more effective. In MA, however, the challenges are greater because the field is broader and more diverse in terms of methodology. Thus, the clarity and consistency of paratexts reflect the perceived quality of MA research, as highlighted by Venturini et al. (2022).

In high-impact journals, textual signals adhere to stricter formulation and clarity standards, thereby reinforcing their value in scientific communication (Del Gesso & Lodhi, 2025). Thus, we treat title↔abstract CC as the intensity of the paratextual signal conditioned by substantive controls (year, country, and SJR tier), without making any causal claims. For analytical purposes, we adopt the convention of high, medium, and low CC and postulate the following:

H_{3a}: articles with high CC between the title↔abstract receive a higher number of citations.

H_{3b}: articles with medium or low CC between the title↔abstract have fewer citations.

3. METHODOLOGICAL PROCEDURES

This research adopts a quantitative approach and a descriptive-documentary design combined with bibliometric techniques. We use NLP to analyze large volumes of texts in a reproducible and scalable manner; here, NLP refers to the use of computational models to transform language into comparable representations, enabling standardization and large-scale comparison (Shchyrba et al., 2024). The incorporation of AI into this type of analysis constitutes a methodological advance in the field (Abbas, 2025).

The methodological design was structured into a pipeline of five macro-stages: (i) data curation; (ii) text preprocessing; (iii) construction of a thematic dictionary and categorization of MA subfields; (iv) CC measurement; and (v) statistical analysis using ordinary least squares (OLS) linear regression models.

Data curation removed duplicates, incomplete records (without an abstract or keywords), and sensitive author-identification metadata (e.g., author names), as these do not contribute to CC modeling and in order to preserve analytical neutrality.

Preprocessing includes text normalization (lowercase equivalences and noise cleaning: whitespace, punctuation, etc.) and stemming (reduction of a word to its root) to mitigate morphological variations and increase comparability across texts (Shchyrba et al., 2024).

Dictionary construction standardizes thematic classification. Based on normalized keywords, we created a controlled vocabulary that associates descriptors with MA subfields according to terminological relevance. Ambiguous cases underwent technical review, and terms with a similarity score below 0.30 were assigned to the residual category, “unclassified.” We defined this threshold *ex ante* and calibrated it empirically based on the similarity distribution (Sentence-BERT [SBERT] + cosine) to reduce lexical ambiguity and preserve the stability of the assignments. This threshold proved to be optimal for this sample and was adopted to ensure the reproducibility of the pipeline.

We deterministically assigned the article’s subfield based on the weighted majority of the descriptors. Using controlled vocabularies and verifying consistency between the title, abstract, and keywords are standard practices in thematic representation studies (Santos et al., 2019).

CC is measured as semantic proximity between the title↔abstract via embeddings, which are dense vector representations that capture the meaning of words, terms, and phrases. These embeddings are processed using SBERT (Paraphrase-MiniLM-L12-v2), an efficient transformer model for textual similarity tasks (Reimers & Gurevych, 2019). The continuous score ranges from 0 to 1 and is derived from the cosine similarity between title↔abstract vectors; values close to 1 indicate greater semantic alignment (Nguyen et al., 2019). For analytical purposes, we classified CC as high (≥ 0.80), medium ($0.60 < 0.80$), or low (< 0.60). These categories were defined *ex ante* based on the sample distribution. The statistical analysis uses OLS regression, which is appropriate for hypothesis testing.

We automated the procedures in Python to enhance reproducibility. The pipeline version, source code, and documentation are available in the public repository https://github.com/renatogo-sc/pln_contabilidade_gerencia, in accordance with the principles of open science (Qiu et al., 2023).

3.1 Sample

The sample was extracted from the Scopus database using the search terms “management accounting” or “managerial accounting” in the title, abstract, and keywords fields. The search covered the period from 1995 to April 2025. The corpus covers business, management, accounting, economics, econometrics, and finance. These fields accounted for 67.3% of the records. The remaining 32.7% were distributed across peripheral fields, such as decision sciences, social sciences, and psychology. We included only documents classified as “article” (77.94% of the total).

Curation procedures included removing duplicates (by DOI/title), excluding incomplete records (without an abstract and/or keywords), and processing metadata. For the purposes of this study, “sensitive data” refers to any metadata that could be used to identify an individual, including personally identifiable information such as author name, email, ORCID, author ID, and profile links. These fields were not used in the modeling and were excluded prior to semantic processing. We did not make inferences based on authors or institutions. For country control, we retained only the first author’s affiliation

(an aggregated variable). This approach provides an objective criterion in cases of multiple affiliations, avoids arbitrariness in weighting multicenter teams, reduces ambiguity and the risk of reidentification, and is sufficient for control purposes since we do not estimate causal effects by country. Instead, we adjust for baseline linguistic and contextual variations. After exclusions, the final portfolio totaled 13,614 articles, which were used in subsequent stages of preprocessing, thematic classification, and CC measurement.

3.2 Variables

The variables were organized into dependent, independent, and control variables.

3.2.1 Dependent

(i) Coherence (Score): A score of semantic similarity between the title↔abstract on a continuous scale from 0 to 1. This score was adapted to measure CC in this study (Reimers & Gurevych, 2019) and was used as an outcome in H_{1a} and H_{1b}. (ii) Citations: Log(1 + cited by). This is an indicator of academic visibility (Friske et al., 2023; Shchyrba et al., 2024) and was used in H_{3a} and H_{3b}.

3.2.2 Independent

(iii) Coherence classification: Discretization of the Coherence variable (Score) into three ex ante ranges: high (≥ 0.80), medium ($0.60 < 0.80$), and low (< 0.60). These ranges were defined based on the sample distribution to compare alignment levels in H_{3a} and H_{3b}. This operationalization is an original methodological contribution of this study. (iv) Thematic group: A subfield assigned by a controlled dictionary of keywords (curated vocabulary) with deterministic assignment based on the weighted majority of descriptors. This is also a methodological contribution used in H_{1a} and H_{1b}. (v) best quartile (Qiu et al., 2023): An aggregated SJR tier used to test Q1–Q2 (higher impact) versus Q3–Q4 (lower impact). It is used as a proxy for editorial impact in H_{2a} and H_{2b}.

3.2.3 Controls

(vi) Year (Abbas, 2025): Controls for temporal trends in CC and citations. (vii) Country (Babaian & Xu, 2024): An aggregated variable used solely for contextual adjustment. In cases of multiple affiliations, the label is assigned based on the first author's affiliation as a reproducible criterion. It is used exclusively as a geographic-linguistic proxy. We do not infer individual author attributes (e.g., residence, length of stay, or country of origin), nor do we consider the effects of language or “grammatical/lexical richness” on CC. When comparisons are presented, they are descriptive in nature.

To capture this specific aspect of textual production, we define CC as a measurable property of the title↔abstract pair. This pair functions as an instrumental indicator that reinforces accountability in scientific communication. It also monitors communication patterns within the field independently of conceptual disputes. CC is necessary but not sufficient for critical education, which also depends on pedagogical mediation, theoretical plurality, and reflective practices.

3.3 Semantic Processing and Statistical Tests

3.3.1 Formation of categories

The eight thematic categories – budgeting, costs, MCS, performance evaluation, organizational behavior, accounting education, other topics, and unclassified – are derived from syntheses of the field and conceptual frameworks of MA presented in Section 2.1. This framework was chosen because it (i) covers families of recurring problems in the literature and teaching of the field, (ii) minimizes semantic overlap among labels, and (iii) preserves historical comparability. The “other topics” category encompasses interdisciplinary or emerging topics with dispersed terminology (e.g., ESG, innovation, governance, digital transformation) when no single subfield is unequivocally predominant. It serves as a neutral analytical category. “Unclassified,” in turn, is a technical residual that includes terms whose similarity matching (SBERT + cosine) fell below the 0.30 threshold and/or articles that do not achieve predominance even after applying the weighted majority of descriptors (requiring ≥ 2 valid matches). This class is not subject to substantive review, does not support theoretical inferences, and is used solely for diagnostic and control purposes within the model.

3.3.2 Lexical dictionary

Using Merchant’s (2010) framework as a guide, the study began with 68,375 normalized keywords. Filters were applied to remove isolated numbers, non-informative acronyms, and excessively long expressions (Pitarch et al., 2023). This process resulted in 26,125 common terms that were mapped by semantic similarity to the categories. This approach followed the best practices for controlled vocabularies (Santos et al., 2019). Ambiguous cases, such as “performance” (which may refer to performance evaluation or MCS), “governance” (which may fall under other topics or MCS), and “learning” (which may indicate accounting education or organizational behavior), as well as acronyms like “ABC,” were addressed using a decision protocol that required: (i) co-occurrence with diagnostic descriptors from the category’s vocabulary, (ii) a weighted majority of the article’s descriptors (classification only with ≥ 2 valid matches), and (iii) a technical review when similarity fell within the gray zone (0.30–0.40). Terms with a similarity of less than 0.30 (SBERT + cosine) were not assigned and constitute the “unclassified” category (technical residual without substantive review). Ultimately, 25,816 terms were classified, while 309 remained unclassified. Each category comprises a set of terms representing the analyzed scientific publication (see Appendix).

3.3.3 Inverted index and article assignment rule

To reduce lexical density biases, where categories with larger dictionaries tend to “win” by volume, we implemented an inverted index. For each term in the dictionary, we maintain a list of articles in which the term occurs, which is a typical retrieval structure. Then, we calculate a normalized score for each article i and category k . The score is given by $score_{ik} = \frac{\sum_{t \in V_k} 1(t \in i)}{|V_k|}$, where V_k is the vocabulary of category k , $1(t \in i)$ is a function that returns 1 if the term t is present in article i (and 0 otherwise), and $|V_k|$ is the total size of the vocabulary of category k . This normalization is essential because it weights the occurrence by the size of the category’s vocabulary, which prevents larger vocabularies from dominating the ranking. For example, if an article contains three terms from a vocabulary of 10, the initial score would be $3/10 = 0.3$. The subcategory of article i is

defined by the weighted majority of its descriptors. For robustness, we require a minimum of two valid matches with the vocabulary (Araque et al., 2019).

3.3.4 Measuring CC (title↔abstract)

Coherence (Score) is the cosine similarity between the sentence-level embeddings of the title and abstract generated by SBERT (Paraphrase-MiniLM-L12-v2), a lightweight transformer model calibrated for textual similarity in short texts. The score ranges from 0 to 1, with higher values indicating greater semantic alignment. Scores in the bottom 1% were manually inspected to ensure consistency (Nguyen et al., 2019; Reimers & Gurevych, 2019).

3.3.5 Analytical use and testing

To interpret H₃, the continuous score was pre-discretized into three categories: high (≥ 0.80), medium ($0.60 < 0.80$), and low (< 0.60). This categorization was based on the empirical distribution of the sample and interpretability criteria. The cutoff points were set to ensure reproducibility of the pipeline.

4. ANALYSIS OF RESULTS

We analyzed CC in the corpus of MA prior to test H₁, H₂ and H₃. In line with the objective, we report the following: (H₁) variation in CC by subfield, (H₂) differences in CC by SJR editorial tier, and (H₃) an association between CC levels and citations conditional on the model's controls.

Our analysis of title↔abstract similarity scores revealed a predominance of medium CC in the original sample ($n = 13,614$). Figure 2 shows a left-skewed distribution with the highest concentration between 0.60 and 0.80 ($n = 8,573$, or 62.9%), followed by low CC ($n = 3,862$, or 28.4%) and high CC ($n = 1,179$, or 8.7%).

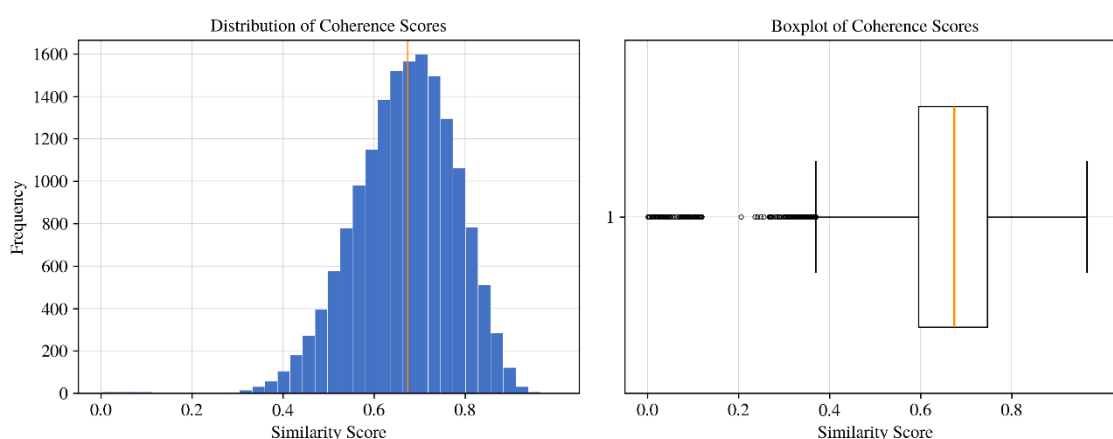


Figure 2 Distribution and dispersion of title↔abstract similarity scores

Source: Prepared by the authors.

The median (orange line), close to 0.70, and the interquartile range ($Q1 \approx 0.45$; $Q3 \approx 0.75$), suggest a relatively consistent pattern of coherence and indicate moderate dispersion. The long left tail, marked by outliers with scores close to zero, indicates critical cases of misalignment between the title↔abstract ($n = 136$). These cases were subsequently inspected manually to ensure the robustness of the process. For illustrative

purposes, consider a title such as “Cost Management in Family-Owned Microenterprises,” accompanied by an abstract describing “an experimental study on game-based learning for accounting students.” This constitutes a critical misalignment. In this case, anticipating the content based solely on the title tends to be less informative, which characterizes a semantic disconnect captured by the metric. Higher CC implies less cognitive effort is required to understand the topic, while lower CC indicates a mismatch between the title’s promise and the abstract’s delivery. In thematic representation studies, misalignments between the title↔abstract may reflect errors, discursive choices, and institutional contexts. This is in line with Devitt’s (2015) distinction between genre competence (mastery of conventions) and genre performance. Misalignments also reduce the quality of representation and the retrievability of information. This motivates the objective measurement of coherence (Santos et al., 2019).

Within the MiniLM architecture, the SBERT model has proven efficient in terms of the performance-computational cost trade-off reported in the literature (Cer et al., 2018; Devlin et al., 2019; Reimers & Gurevych, 2019).

4.1 Testing Hypotheses H_{1a} and H_{1b} : CC by Thematic Subfield

Categorical variables require a reference contrast for identifiability in regression models with an intercept (OLS). We adopted the “budgeting” thematic group because it is the most frequent subarea in the sample ($n = 3,485$; 25.6%), which improves precision. This choice does not affect the overall tests of H_{1a} and H_{1b} , only the interpretation of the relative coefficients.

Table 1*OLS regression results for H_{1a} and H_{1b}*

Thematic group	Coefficient	SE	t	p-value	95% CI lower	95% CI upper
Performance evaluation	-0.0035	0.0061	-0.5686	0.5697	-0.0154	0.0085
Organizational behavior	-0.0218	0.0051	-4.2418	0***	-0.0318	-0.0117
Cost	-0.0233	0.0047	-4.9910	0***	-0.0324	-0.0141
Accounting education	-0.0021	0.0074	-0.2762	0.7824	-0.0166	0.0125
Other topics	0.0027	0.0054	0.4996	0.6174	-0.0079	0.0132
MCS	-0.0206	0.0053	-3.8808	0.0001***	-0.0309	-0.0102
Unclassified	-0.0095	0.0043	-2.2137	0.0269*	-0.0179	-0.0011
Control variables						
Intercept	-1.3833	0.5001	-2.7659	0.0057**	-2.3637	-0.4029
Year	0.001	0.0002	4.1473	0***	0.0005	0.0015
Quartile	-0.0099	0.0015	-6.3900	0***	-0.0129	-0.0068

*SE = standard error; 95% CI = 95% confidence interval; OLS = ordinary least squares.*** = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$.**Source: Prepared by the authors.*

The analysis across thematic groups revealed that organizational behavior (-0.0218; $p < 0.001$), costs (-0.0233; $p < 0.001$), and SCG (-0.0206; $p = 0.0001$) had significantly lower coherence scores than the reference group. This suggests that these subfields require more effort from readers to interpret, possibly because they address topics with less rhetorical standardization or greater epistemic complexity. This finding is supported by relevance theory (Wilson & Sperber, 2004).

The unclassified thematic group (-0.0095; $p = 0.0269$) indicates a small effect size, which is consistent with the heterogeneity of the residual group. In contrast, the performance evaluation (-0.0035; $p = 0.5697$), accounting education (-0.0021; $p = 0.7824$), and other topics (0.0027; $p = 0.6174$) coefficients were not significant. This indicates that CC does not differ from the reference category for these groups. Thus, H_{1a} is partially supported: the reference group (budgeting) exhibits greater relative coherence as an established subfield. However, H_{1b} is partially rejected since not all maturing subfields exhibited lower coherence. This balance stems from maximizing the positive cognitive effect, or the gain of relevant information, combined with pursuing the path of least resistance, which reduces the reader's interpretive burden. Therefore, the higher the CC, the lower the cognitive cost, and the greater the informational efficiency of the scientific text.

Additionally, we analyzed the distribution of CC over time and among the top 20 countries in the sample. The annual coefficient of 0.001 indicates a consistent increase in the average CC over the years, which may reflect a maturation of editorial practices or the spread of more uniform rhetorical patterns. China (average 0.71) and South Korea (0.68), both countries with high scientific output, exhibited higher CC. India (0.61), Sweden (0.62), and Ukraine (0.63) recorded the lowest indices. Brazil (0.66) ranked slightly above the overall average, indicating minor differences among countries. The details of this analysis are available in the project's online repository, which features an interactive map showing distributions by country and year.

In light of relevance theory, we interpret differences in CC as variations in the balance between cognitive effect and processing effort. In subfields with negative coefficients, titles tend to signal less (lower anticipatory effect), and abstracts tend to reflect the informational core less concisely, requiring more effort to understand the same information. In subfields with no statistical difference, the title↔abstract pair preserves sufficient alignment for the same core meaning. The reading remains strictly textual, without a paradigmatic evaluation of the subfields. Evidence regarding citations should be interpreted as an average association (a proxy for impact or visibility), not an epistemological evaluation of the field.

Table 2

General statistics of the OLS model for H_{1a} and H_{1b}

Statistic	Value	Description
R^2	0.071	Proportion of variance explained (acceptable in scientometric studies)
Adjusted R^2	0.054	R^2 adjusted for variables (low variance, but typical in broad studies)
F-statistic	4.176	Model significance test (a high value indicates good fit)
Prob (F-statistic)	9.82e-46	p-value of F (< 0.05 confirms the overall significance of the model)
Log-likelihood	5,080.6	A higher value indicates better fit

AIC	-9,919	A lower value indicates better fit; useful for comparison
-----	--------	---

AIC = Akaike information criterion; OLS = ordinary least squares.

Source: Prepared by the authors.

The general statistics of the OLS model for H_{1a} and H_{1b} indicate robustness. The adjusted R² (0.054), which is typical in scientometric studies, reflects modest variance explanation. Meanwhile, the F-statistic (4.176) has a p-value of 9.82e-46 and confirms the model's overall significance, suggesting that the independent variables have combined explanatory power. The log-likelihood (5,080.6) and the Akaike information criterion (AIC) (-9,919) support the model's fit and are useful for comparison with other models. To assess limitations, residual diagnostics are examined.

Table 3

Diagnostics of the OLS model for H_{1a} and H_{1b}

Statistic	Value	Description
Omnibus	565.483	Normality test of the residuals
Prob (Omnibus)	0.000	p-value of the Omnibus (< 0.05 indicates non-normal residuals)
Durbin-Watson	1.958	Autocorrelation of the residuals (close to 2, absence of autocorrelation)
JB	756.090	Normality test of the residuals
Prob (JB)	6.56e-165	p-value of the JB (< 0.05 indicates non-normal residuals)
Cond. No.	1.75e+06	Condition number (a high value suggests possible multicollinearity)

JB = Jarque-Bera; OLS = ordinary least squares.

Source: Prepared by the authors.

The OLS model for H_{1a} and H_{1b} is robust and interpretable. However, the normality tests (Omnibus: 565.483, p = 0.000; Jarque-Bera: 756.090, p = 6.56e-165) indicate non-normal residuals, which are common in large samples. Additionally, the high condition number (1.75e+06) suggests multicollinearity, which may be due to the large number of country categories and could affect stability. The Durbin-Watson statistic (1.958), which is close to 2, confirms the absence of autocorrelation. Variables such as year (significant positive coefficient) and best quartile (significant negative coefficient) also reinforce the model's validity.

4.2 Testing Hypotheses H_{2a} and H_{2b}: CC by Journal Impact

The Q3–Q4 group was chosen as the reference category because it represents the tier with the lowest impact in the sample. This allows us to test whether more prestigious journals (Q1–Q2) tend to have higher CC.

Table 4*OLS regression results for H_{2a} and H_{2b}*

Dependent variable	Coefficient	SE	t	p-value	95% CI lower	95% CI upper
Coherence (Score)						
(Q1-Q2)	0.0207	0.0035	5.9960	0.0000***	0.0140	0.0275
(Q3-Q4)	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000
Control variables						
Intercept	-1.2666	0.4988	-2.5394	0.0111*	-2.2444	-0.2889
Year	0.0009	0.0002	3.7948	0.0001***	0.0004	0.0014

*SE = standard error; 95% CI = 95% confidence interval; OLS = ordinary least squares.*** = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$.**Source: Prepared by the authors.*

Analysis of the coefficients indicates that, on average, articles published in Q1–Q2 journals have a coherence score that is 0.0207 units higher than the reference category (Q3–Q4), which is a highly significant result ($p < 0.001$). This higher mean score is consistent with the editorial practices of Q1–Q2 journals, which require clarity and standardization of initial components such as the title, abstract, and keywords. These practices increase alignment between the title↔abstract (Garcia et al., 2019). This finding provides empirical support for H_{2a}, suggesting that higher-impact journals promote more uniform and demanding standards of textual construction. Consequently, H_{2b} is also confirmed since the group with lower impact exhibits significantly lower coherence scores. Furthermore, the positive and significant coefficient of the year variable (0.0009; $p = 0.0001$) indicates continuous improvement in CC over time. This trend may be associated with the consolidation of editorial guidelines, the more widespread use of writing support tools, and the maturation of the scientific community in terms of style and clarity of argumentation.

A descriptive analysis of CC scores by country and year among the top 20 reinforces these patterns: China (average 0.71) and Germany (0.67) had the highest scores, while India (0.61) and Sweden (0.62) had the lowest, with Brazil (0.66) ranking above average. An interactive visualization of this distribution is available in the project's online repository.

In light of text genre theory, the contrast between Q1–Q2 and Q3–Q4 aligns with the idea that more restrictive editorial guidelines and established genre conventions (e.g., an informative title and an abstract that covers the expected elements, such as background, gap, objective, method, findings, and implications) influence the relationship between the title↔abstract, increasing mirroring and, consequently, CC. The result is associative, with an average effect that is positive and statistically significant in this case, and is strictly textual. It does not imply a judgment of scientific “quality.”

Table 5

General statistics of the OLS model for H_{2a} and H_{2b}

Statistic	Value	Description
R ²	0.063	Proportion of variance explained (acceptable in scientometric studies)
Adjusted R ²	0.047	R ² adjusted for variables (low variance, but typical in broad studies)
F-statistic	3.932	Model significance test (a high value indicates good fit)
Prob (F-statistic)	4.44e-39	p-value of F (< 0.05 confirms the overall significance of the model)
Log-likelihood	5,053.4	A higher value indicates better fit
AIC	-9,879	A lower value indicates better fit; useful for comparison

AIC = Akaike information criterion; OLS = ordinary least squares.

Source: Prepared by the authors.

Similar to the analysis for H_{1a} and H_{1b}, the overall statistics of the OLS model for H_{2a} and H_{2b} indicate robustness. The adjusted R² (0.047), which is typical in scientometric studies, reflects a modest explanation of variance. Meanwhile, the F-statistic (3.932), which has a p-value of 4.44e-39, confirms the model's overall significance. This suggests that the independent variables have combined explanatory power. The log-likelihood

(5,053.4) and AIC (-9,879) support the model's fit and are useful for comparison with other models. To assess limitations, we examine the residuals.

Table 6

Diagnostics of the OLS model for H_{2a} and H_{2b}

Statistic	Value	Description
Omnibus	579.289	Normality test of the residuals
Prob (Omnibus)	0.000	p-value of the Omnibus (< 0.05 indicates non-normal residuals)
Durbin-Watson	1.954	Autocorrelation of the residuals (close to 2, absence of autocorrelation)
JB	777.266	Normality test of the residuals
Prob (JB)	1.66e-169	p-value of the JB (< 0.05 indicates non-normal residuals)
Cond. No.	1.75e+06	Condition number (a high value suggests possible multicollinearity)

JB = Jarque-Bera; OLS = ordinary least squares.

Source: Prepared by the authors.

Similar to the analysis for H_{1a} and H_{1b}, the OLS model for H_{2a} and H_{2b} is robust and interpretable, even though the normality tests (Omnibus: 579.289, $p = 0.000$; JB: 777.266, $p = 1.66e-169$) indicate non-normal residuals, which is common in large samples. Additionally, the high condition number (1.75e+06) suggests multicollinearity, possibly due to the many country categories, which affects stability. The Durbin-Watson statistic (1.954), which is close to 2, confirms the absence of autocorrelation. Variables such as year (with a significant positive coefficient) and the tier with the greatest impact (Q1–Q2, with a significant positive coefficient) reinforce the validity of the model.

4.3 Testing Hypotheses H_{3a} and H_{3b}: Association between CC and Citations

The reference category comprises articles with high coherence (score ≥ 0.80). We contrast the other levels against this category in our analyses. This approach allows us to identify high-impact patterns and offers valuable insights into how coherence between the title↔abstract may influence academic impact. This is especially relevant in fields such as MA, which is still consolidating its discursive standards. Observing the significant association between high CC and a higher number of citations suggests that greater textual clarity may facilitate engagement with and visibility of the article. This creates a practical reference for authors in the field.

Table 7*OLS regression results for H_{3a} and H_{3b}*

Variables	Coefficient	SE	z	p-value	95% CI lower	95% CI upper
Dependent: log(1 + cited by)						
Independent: mean (0.60–<0.80)	-0.0599	0.0359	-1.6698	0.095	-0.1302	0.0104
Independent: low (<0.60)	-0.2019	0.0382	-5.2847	0.000***	-0.2768	-0.1270
Control variables						
Intercept	157.1231	3.3138	47.4152	0.000***	150.6282	163.6180
Year	-0.0770	0.0016	-46.9122	0.000***	-0.0802	-0.0738
Q1	1.0779	0.0285	37.7838	0.000***	1.0220	1.1339
Q2	0.2433	0.0301	8.0831	0.000***	0.1843	0.3023
Q3	-0.1409	0.0344	-4.0997	0.000***	-0.2083	-0.0736
Q4	-0.6576	0.0347	-18.9552	0.000***	-0.7256	-0.5896

*SE = standard error (robust, HC3); 95% CI = 95% confidence interval; OLS = ordinary least squares.*** = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$.**Source: Prepared by the authors.*

Analysis of the coefficients reveals that articles with low CC have a $\log(1 + \text{number of citations})$ that is 0.2019 units lower than articles with high CC. This is a statistically robust result. It suggests that to improve the visibility and academic impact of articles in MA, authors should prioritize textual clarity, particularly in aligning the title↔abstract. This recommendation aligns with the idea that cohesive and accessible communication facilitates the screening process and increases the likelihood of being cited. Furthermore, this finding reinforces the importance of using tools, such as the dictionary provided in the Appendix, to reduce ambiguity and facilitate language standardization across subfields. However, articles with a medium CC show a negative coefficient of -0.0599, which is not significant at the 5% level ($p = 0.095$). Therefore, a conclusive inference cannot be made.

These results confirm H_{3a} , which states that high CC is associated with greater academic visibility. The results also partially support H_{3b} , since only low CC had a significant negative impact while medium CC did not differ substantially from high CC. In this sense, the dictionary developed for MA is a practical tool for authors and editors. By categorizing terms according to subfields, it reduces ambiguity and standardizes communication, which is especially useful in a field that faces challenges regarding clarity and consistency, as discussed by Merchant (2010).

Regarding the reference category, low CC reduces $\log(1 + \text{citations})$ by -0.2019 ($p < 0.001$), and medium CC has an effect of -0.0599 ($p = 0.095$). These results provide partial support for the hypotheses. The empirical pattern in which articles with high CC systematically exhibit higher $\log(1 + \text{citations})$ than other levels constitutes statistical evidence that higher CC is associated with greater impact, as measured by citations. The association between high CC and citations aligns with signaling theory, which suggests that consistent paratexts (e.g., title↔abstract alignment) reduce interpretation costs and increase retrievability, thereby impacting visibility (Garcia et al., 2019).

The control variables strengthen the empirical validity. The quartiles (Q1: 1.0779, $p < 0.001$; Q4: -0.6576, $p < 0.001$) reflect editorial prestige. Year (-0.0770, $p < 0.001$) indicates a slight decline in citations over time, possibly due to obsolescence. The intercept (157.1231, $p < 0.001$) captures the baseline effect.

Table 8

General statistics of the OLS model for H_{3a} and H_{3b}

Statistic	Value	Description
R^2	0.362	Proportion of variance explained (acceptable in scientometric studies)
Adjusted R^2	0.362	R^2 adjusted for variables (low variance, but typical in broad studies)
F-statistic	1212	Model significance test (a high value indicates good fit)
Prob (F-statistic)	0.000	p-value of F (< 0.05 confirms the overall significance of the model)
Log-likelihood	-20,973	A higher value indicates better fit
AIC	4.196e+04	A lower value indicates better fit; useful for comparison

AIC = Akaike information criterion; OLS = ordinary least squares.

Source: Prepared by the authors.

The general statistics show that R^2 (0.362) explains 36.2% of the variance, which is consistent with previous scientometric studies. The identical adjusted R^2 (0.362) reinforces the stability of the fit. The F-statistic (1212) has a p-value of less than 0.001, which confirms the model's overall significance and suggests that the independent

variables have strong combined explanatory power. The log likelihood (-20.973) and AIC (4.196e+04) support the model's fit and are useful for comparing it to other models.

Table 9

Diagnostics of the OLS model for H_{3a} and H_{3b}

Statistic	Value	Description
Omnibus	96.084	Normality test of the residuals
Prob (Omnibus)	0.000	p-value of the Omnibus (< 0.05 indicates non-normal residuals)
Durbin-Watson	1.592	Autocorrelation of the residuals (close to 2, absence of autocorrelation)
JB	72.628	Normality test of the residuals
Prob (JB)	1.69e-16	p-value of the JB (< 0.05 indicates non-normal residuals)
Cond. No.	6.00e+05	Condition number (a high value suggests possible multicollinearity)

JB = Jarque-Bera; OLS = ordinary least squares.

Source: Prepared by the authors.

The OLS model for H_{3a} and H_{3b} is interpretable, despite the non-normality of the residuals (Omnibus: 96.084, $p = 0.000$; JB: 72.628, $p = 1.69\text{e-}16$), which is common in large samples, and the high Cond. No. (6.00e+05), suggesting multicollinearity, possibly due to country categories. The Durbin-Watson statistic (1.592) is slightly below 2 and indicates mild autocorrelation. However, it does not compromise the model's overall validity, although caution should still be exercised when interpreting the results.

5. CONCLUSIONS AND RECOMMENDATIONS

This study measured the coherence (CC) between the title↔abstract of 13,614 MA articles published between 1995 and April 2025. Due to the conceptual-operational gap, or the lack of consensus on how to measure and link CC to impact metrics in a comparable way, we applied large-scale NLP techniques. Our findings revealed that the majority of articles exhibit medium (62.9%) or low (28.4%) levels of CC, while only a small percentage (8.7%) reach high levels. These results reveal persistent tensions between discursive clarity and conceptual density, which may affect readability, interpretability, and possibly the circulation of scientific knowledge.

Statistical analysis showed that well-established subfields, such as budgeting, tend to be more coherent. In contrast, areas such as costs and MCS face greater challenges in textual construction, suggesting variations in rhetorical standardization or greater epistemic complexity. Additionally, articles published in high-impact journals (Q1–Q2) exhibited higher CC, reinforcing the hypothesis that more selective journals impose stricter linguistic and structural standards.

Hypotheses H_{1a} and H_{1b} were partially confirmed, while H_{2a}, H_{2b} and H_{3a} found robust empirical support. Hypothesis H_{3b} was partially confirmed since the difference between articles with medium and high CC was not statistically significant. This nuance suggests that although extreme coherence is desirable, marginal gains in impact may plateau after a certain threshold. Taken together, our quantitative results in MA are consistent with qualitative evidence from related fields regarding the role of titles, abstracts, and keywords in retrievability and visibility (Garcia et al., 2019; Pereira, 2013; Santos et al., 2019).

The theoretical implications of this study are significant in three ways. First, it reinforces relevance theory (Wilson & Sperber, 2004) by showing that the degree of compatibility between the title↔abstract affects the cognitive effort required for comprehension. Second, the study contributes to research on scientific text genres (Swales, 1990) by showing how rhetorical patterns vary according to journal type and subfield maturity. Third, the study advances the application of signaling theory (Spence, 1973) by showing that the structure of a text can influence how an article is received and its impact, regardless of its technical content.

From a practical standpoint, the results offer diagnostic and analytical insights with implications for editorial processes, criteria for evaluating scientific output, and student training strategies without being prescriptive. The connection to SDG 4 (Quality Education) transcends the technical scope, pointing toward more socially engaged scientific practices. CC serves as a tool for teaching and editorial governance, providing an observable indicator of paratextual signals. However, it does not directly measure cost nor guarantee impact. Its utility is instrumental – providing transparency, curation, and screening – and depends on complementary editorial and pedagogical practices.

From a methodological standpoint, the study contributes a replicable protocol consisting of five automated macro steps in Python operated via VLAB@UFSC, a high-performance scientific computing environment. This technical setup enabled the complete processing of the sample in three minutes and three seconds with precision and traceability, setting a model for open and reproducible science. The scripts, generated dictionary, interactive maps, and processed data are available in a public GitHub repository, aligning with the principles of the third mission of universities and commitment to open science (Compagnucci & Spigarelli, 2020; Qiu et al., 2023; Russo, 2014).

Despite the robustness of the models, technical and epistemological limitations were observed. Non-normal residuals and multicollinearity among variables such as year and quartile indicate possible structural noise, requiring caution when drawing causal inferences. Approximately 19.8% of the sample ($n = 2,691$) could not be classified thematically, primarily due to the absence of relevant metadata fields, highlighting the dependence of the automated approach on database quality. Additionally, setting the threshold at less than 0.30 limits the models' explanatory power and restricts direct comparisons across fields of knowledge. We acknowledge that, although the geographic-linguistic proxy used to control for the first author's country is an objective and replicable criterion, it may generate residual heterogeneity that is not fully accounted for by this control. Finally, the modeling does not incorporate information on authorship, academic reputation, or the authors' background. Such factors may influence reading and citation patterns. This methodological choice limits the scope of the study and constitutes an additional epistemological limitation. Similarly, the study does not measure readers' perceptions, which remain outside the scope of the coherence metric. The findings are limited to the textual level. We do not assess the ontological or paradigmatic adequacy of the subfields, nor do we evaluate the intrinsic merit of the field.

Future research could model co-authorship networks, multiple affiliations, and country-versus-quartile interactions in order to more accurately estimate the effects of transactions on citations and CC. Longitudinal analyses could elucidate discursive transformations over time. Incorporating contextual variables, such as language, academic tradition, and institutional profile, may explain geographical differences that current models do not capture. Additionally, refining the dictionary (see the Appendix) using unsupervised techniques (e.g., semantic clustering) inspired by Merchant (2010) ideas may reveal new axes of meaning in the accounting literature. Finally, qualitative

analyses of articles with low coherence can complement the quantitative approach, deepening our understanding of factors that compromise textual clarity. To mitigate the effects of multicollinearity, we suggest the following: (i) dimensionality reduction via principal component analysis (PCA) or partial least squares to generate orthogonal predictors, (ii) penalized regressions (ridge, Lasso, elastic net) for shrinkage and selection under multicollinearity, and (iii) robustness checks involving variance inflation factor diagnostics and alternative specifications to achieve greater stability of the coefficients.

This study establishes a methodological framework for communicative analysis in accounting research by integrating three theoretical frameworks – relevance, text genres, and signaling – into an automated and scalable protocol. The results are expected to provide diagnostic insights that can inform editorial policies, faculty training initiatives, and institutional strategies to promote a rigorous, transparent, and socially responsible scientific culture.

REFERENCES

- Abbas, K. (2025). Management accounting and artificial intelligence: A comprehensive literature review and recommendations for future research. *British Accounting Review*. <https://doi.org/10.1016/j.bar.2025.101551>
- Abernethy, M. A., & Brownell, P. (1999). The role of budgets in organizations facing strategic change: an exploratory study. *Accounting, Organizations and Society*, 24(3), 189-204. [https://doi.org/10.1016/S0361-3682\(98\)00059-2](https://doi.org/10.1016/S0361-3682(98)00059-2)
- Aboagye-Otchere, F., Agyenim-Boateng, C., Enusah, A., & Aryee, T. E. (2021). A review of big data research in accounting. *Intelligent Systems in Accounting, Finance and Management*, 28(4), 268-283. <https://doi.org/10.1002/isaf.1504>
- Alawattage, C., & Wickramasinghe, D. (2024). Teaching strategic management accounting with sustainability. *Accounting Education*, 1-33. <https://doi.org/10.1080/09639284.2024.2404923>
- Ansari, S. L. (1977). An integrated approach to control system design. In *Organizations and Society* (Vol. 2, Number 2). Pergamon Press.
- Araque, O., Zhu, G., & Iglesias, C. A. (2019). A semantic similarity-based perspective of affect lexicons for sentiment analysis. *Knowledge-Based Systems*, 165, 346-359. <https://doi.org/10.1016/j.knosys.2018.12.005>
- Babaian, T., & Xu, J. (2024). Entity recognition from colloquial text. *Decision Support Systems*, 179. <https://doi.org/10.1016/j.dss.2024.114172>
- Barreto, A., Gomes, P., Quesado, P., & O'Sullivan, S. (2024). Management accounting and digital technologies: A science mapping review. In *Proceedings of the 6th International Conference on Advanced Research Methods and Analytics (CARMA 2024)*. Universitat Politècnica de València. <https://doi.org/10.4995/carma2024.2024.17814>
- Becher, T., & Trowler, P. (2001). *Academic tribes and territories: Intellectual enquiry and the culture of disciplines*. Open University Press.
- Beuren, I. M., & Souza, H. R. de. (2025). Theoretical anchoring of management accounting articles: a systematic review. *Meditari Accountancy Research*, 33(3), 763-790. <https://doi.org/10.1108/MEDAR-05-2024-2468>
- Birnberg, J. G., Luft, J., & Shields, M. D. (2006). Psychology theory in management accounting research. In *Handbooks of Management Accounting Research* (Vol. 1, pp. 113-135). [https://doi.org/10.1016/S1751-3243\(06\)01004-2](https://doi.org/10.1016/S1751-3243(06)01004-2)
- Burton, F. G., Summers, S. L., Wilks, T. J., & Wood, D. A. (2022). Relevance of accounting research (ROAR) scores: Ratings of titles and abstracts by accounting

- professionals. *Accounting Horizons*, 36(2), 7-18.
<https://doi.org/10.2308/HORIZONS-2020-147>
- Cer, D., Yang, Y., Kong, S., Hua, N., Limtiaco, N., John, R. St., Constant, N., Guajardo-Cespedes, M., Yuan, S., Tar, C., Sung, Y.-H., Strope, B., & Kurzweil, R. (2018). Universal sentence encoder. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.
<https://doi.org/10.18653/v1/D18-2029>
- Chapman, C. S. (1997). Reflections on a contingent view of accounting. *Accounting, Organizations and Society*, 22(2), 189-205. [https://doi.org/10.1016/S0361-3682\(97\)00001-9](https://doi.org/10.1016/S0361-3682(97)00001-9)
- Colombo, O. (2021). The use of signals in new-venture financing: A review and research agenda. *Journal of Management*, 47(1), 237-259.
<https://doi.org/10.1177/0149206320911090>
- Compagnucci, L., & Spigarelli, F. (2020). The third mission of the university: A systematic literature review on potentials and constraints. *Technological Forecasting and Social Change*, 161.
<https://doi.org/10.1016/j.techfore.2020.120284>
- Deegan, C. (2002). Introduction: The legitimising effect of social and environmental disclosures – A theoretical foundation. *Accounting, Auditing & Accountability Journal*, 15(3), 282-311. <https://doi.org/10.1108/09513570210435852>
- Del Gesso, C., & Lodhi, R. N. (2025). Theories underlying environmental, social and governance (ESG) disclosure: A systematic review of accounting studies. *Journal of Accounting Literature*, 47(2). <https://doi.org/10.1108/jal-08-2023-0143>
- Devitt, A. J. (2015). Genre performances: John Swales' Genre analysis and rhetorical-linguistic genre studies. *Journal of English for Academic Purposes*, 19, 44-51.
<https://doi.org/10.1016/j.jeap.2015.05.008>
- Devlin, J., Chang, M.-W., Lee, K., Google, K. T., & Language, A. I. (2019). *BERT: Pre-training of deep bidirectional transformers for language understanding*.
<https://github.com/tensorflow/tensor2tensor>
- DiMaggio, P. J., & Powell, W. W. (2005). *A gaiola de ferro revisitada: isomorfismo institucional e racionalidade coletiva nos campos organizacionais*. *RAE-Revista de Administração de Empresas*, 45(2), 74-89.
- Endenich, C., & Trapp, R. (2020). Ethical implications of management accounting and control: A systematic review of the contributions from the Journal of Business Ethics. *Journal of Business Ethics*, 163(2), 309-328.
<https://doi.org/10.1007/s10551-018-4034-8>
- Ensslin, L., Ensslin, S. R., & Pinto, H. M. (2013). Processo de investigação e análise bibliométrica: avaliação da qualidade dos serviços bancários. *Revista de Administração Contemporânea*, 17, 325-349. <https://doi.org/10.1590/S1415-65552013000300005>
- Ferreira, A., & Otley, D. (2009). The design and use of performance management systems: An extended framework for analysis. *Management Accounting Research*, 20(4), 263-282. <https://doi.org/10.1016/j.mar.2009.07.003>
- Friske, W., Hoelscher, S. A., & Nikolov, A. N. (2023). The impact of voluntary sustainability reporting on firm value: Insights from signaling theory. *Journal of the Academy of Marketing Science*, 51(2), 372-392.
<https://doi.org/10.1007/s11747-022-00879-2>
- Garcia, D. C. F., Gattaz, C. C., & Gattaz, N. C. (2019). The relevance of title, abstract and keywords for scientific paper writing. *Revista de Administração Contemporânea*, 23(3). <https://doi.org/10.1590/1982-7849rac2019190178>

- Gray, R. (2010). Is accounting for sustainability actually accounting for sustainability...and how would we know? An exploration of narratives of organisations and the planet. *Accounting, Organizations and Society*, 35(1), 47-62. <https://doi.org/10.1016/j.aos.2009.04.006>
- Henri, J. F. (2006). Management control systems and strategy: A resource-based perspective. *Accounting, Organizations and Society*, 31(6), 529-558. <https://doi.org/10.1016/j.aos.2005.07.001>
- Hyland, K. (2004). Disciplinary interactions: Metadiscourse in L2 postgraduate writing. *Journal of Second Language Writing*, 13(2), 133-151. <https://doi.org/10.1016/j.jslw.2004.02.001>
- Hyland, K., & Tse, P. (2005). Hooking the reader: A corpus study of evaluative that in abstracts. *English for Specific Purposes*, 24(2), 123-139. <https://doi.org/10.1016/j.esp.2004.02.002>
- Ittner, C. D., & Larcker, D. F. (1998). Innovations in performance measurement trends and research implications. *Journal of Management Accounting Research*, 10, 205.
- Jamali, H. R., & Nikzad, M. (2011). Article title type and its relation with the number of downloads and citations. *Scientometrics*, 88(2), 653-661. <https://doi.org/10.1007/s11192-011-0412-z>
- Kloppel, F. V., Lunkes, R. J., & Schmitz, E. (2012). Análise sobre a(s) linha(s) de pesquisa dos professores de programas de pós-graduação em contabilidade. *Revista Capital Científico*, 11(1), 1-17. <https://revistas.unicentro.br/index.php/capitalcientifico/article/view/2067>
- Krishnamoorthy, G., Cohen, J., College, B., & Wright, A. (2004). The corporate governance mosaic and financial reporting quality. *Journal of Accounting Literature*, 23, 87-152. <https://ssrn.com/abstract=1086743>
- Lacey, H. (2011). A imparcialidade da ciência e as responsabilidades dos cientistas. *Scientiae Studia*, 9(3), 487-500. <https://doi.org/10.1590/S1678-31662011000300003>
- Lillis, A. M. (2002). Managing multiple dimensions of manufacturing performance-an exploratory study. *Accounting, Organizations and Society*, 27(6), 497-529. [https://doi.org/10.1016/S0361-3682\(01\)00032-0](https://doi.org/10.1016/S0361-3682(01)00032-0)
- Luft, J., & Shields, M. D. (2003). *Mapping management accounting: Graphics and guidelines for theory-consistent empirical research* [Working Paper]. Social Science Research Network. <https://doi.org/10.2139/ssrn.305959>
- Lunkes, R. J., Ripoll-Feliu, V. M., & Rosa, F. S. da. (2011). Management accounting: A study in Brazilian journals and Spanish language. *Revista de Contabilidade e Organizações*, 5(13). <https://doi.org/10.11606/rco.v5i13.34808>
- Lunkes, R. J., Ripoll-Feliu, V. M., & Rosa, F. S. da. (2012). Autores y citas en contabilidad de gestión en revistas de lengua española. *Revista Venezolana de Gerencia*, 17(58). <https://doi.org/10.31876/revista.v17i58.10700>
- Lunkes, R. J., Rosa, F. S. da, Gasparetto, V., & Balduino, E. (2011). Análise da produção científica e formação de doutores em contabilidade gerencial: um estudo no cenário brasileiro. *Advances in Scientific and Applied Accounting*, 4(3), 361-378. <https://doi.org/10.14392/ASAA.2011040304>
- Malmi, T., & Brown, D. A. (2008). Management control systems as a package – Opportunities, challenges and research directions. *Management Accounting Research*, 19(4), 287-300. <https://doi.org/10.1016/j.mar.2008.09.003>
- Malmi, T., & Granlund, M. (2011). *In search of management accounting theory* [Working Paper]. Social Science Research Network. <https://doi.org/10.2139/ssrn.804004>

- Merchant, K. A. (2007). O modelo do sistema de orçamento corporativo: influências no comportamento e no desempenho gerencial. *Revista de Contabilidade e Organizações*, 1(1). <https://doi.org/10.11606/rco.v1i1.34700>
- Merchant, K. A. (2010). Paradigms in accounting research: A view from North America. *Management Accounting Research*, 21(2), 116-120. <https://doi.org/10.1016/j.mar.2010.02.004>
- Miller, B. W. (2021). Beyond instrumentalism? An instructional project using communicative learning in management accounting. *Journal of Education for Business*, 96(1), 44-51. <https://doi.org/10.1080/08832323.2020.1747379>
- Müller-Stewens, B., Widener, S. K., Möller, K., & Steinmann, J. C. (2020). The role of diagnostic and interactive control uses in innovation. *Accounting, Organizations and Society*, 80. <https://doi.org/10.1016/j.aos.2019.101078>
- Nguyen, H. T., Duong, P. H., & Cambria, E. (2019). Learning short-text semantic similarity with word embedding and external knowledge sources. *Knowledge-Based Systems*, 182. <https://doi.org/10.1016/j.knosys.2019.07.013>
- Otley, D. (1978). Budget use and managerial performance. *Journal of Accounting Research*, 16(1). <https://doi.org/10.2307/2490414>
- Otley, D. (1999). Performance management: A framework for management control systems research. *Management Accounting Research*, 10(4), 363-382. <https://doi.org/10.1006/mare.1999.0115>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Pereira, M. G. (2013). O resumo de um artigo científico. *Epidemiologia e Serviços de Saúde*, 22(4), 707-708. <https://doi.org/10.5123/s1679-49742013000400017>
- Pitarch, L., Bernad, J., Dranca, L., Bobed Lisboa, C., & Gracia, J. (2023). No clues, good clues: Out of context Lexical Relation Classification. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. <https://doi.org/10.18653/v1/2023.acl-long.308>
- Plesner, Å. (2024). Seizing overflows: Exploring how accounting becomes emancipatory. *Qualitative Research in Accounting and Management*, 21(5): 444-464. <https://doi.org/10.1108/QRAM-06-2023-0109>
- Qiu, F., Hu, N., Liang, P., & Dow, K. (2023). Measuring management accounting practices using textual analysis. *Management Accounting Research*, 58. <https://doi.org/10.1016/j.mar.2022.100818>
- Reimers, N., & Gurevych, I. (2019). Sentence embeddings using Siamese BERT-networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. <https://doi.org/10.18653/v1/D19-1410>
- Rodrigue, M., Magnan, M., & Cho, C. H. (2013). Is environmental governance substantive or symbolic? An Empirical Investigation. *Journal of Business Ethics*, 114(1), 107-129. <https://doi.org/10.1007/s10551-012-1331-5>
- Russo, M. (2014). Ética e integridade na ciência: da responsabilidade do cientista à responsabilidade coletiva. *Estudos Avançados*, 28(80). <https://doi.org/10.1590/S0103-40142014000100016>

- Sangster, A. (2015). You cannot judge a book by its cover: The problems with journal rankings. *Accounting Education*, 24(3), 175-186. <https://doi.org/10.1080/09639284.2015.1055929>
- Santos, F. C. L., Mollica, M. C. M., & Guedes, V. L. S. (2019). Coherence in the thematic representation of scientific articles in the public health area. *Perspectivas em Ciência da Informação*, 24(3), 214-233. <https://doi.org/10.1590/1981-5344/3727>
- Savić, B., & Pavlović, V. (2023). Impact of digitalization on the accounting profession. In *Contributions to finance and accounting: Part F233* (pp. 19-34). Springer Nature. https://doi.org/10.1007/978-3-031-23269-5_2
- Scapens, R. W. (2006). Understanding management accounting practices: A personal journey. *British Accounting Review*, 38(1), 1-30. <https://doi.org/10.1016/j.bar.2005.10.002>
- Scarpellini, S., Marín-Vinuesa, L. M., Aranda-Usón, A., & Portillo-Tarragona, P. (2020). Dynamic capabilities and environmental accounting for the circular economy in businesses. *Sustainability Accounting, Management and Policy Journal*, 11(7), 1129-1158. <https://doi.org/10.1108/SAMPJ-04-2019-0150>
- Shchyrba, I., Savitskaya, M., Fursa, T., Yeremian, O., & Ostropolska, Y. (2024). Management accounting: The latest technologies, ChatGPT capabilities. *Financial and Credit Activity: Problems of Theory and Practice*, 1(54), 160-172. <https://doi.org/10.55643/fcaptop.1.54.2024.4307>
- Shields, M. D. (1989). A behavioral model for implementing cost management systems. *Journal of Cost Management*, 3(3), 17-27.
- Shields, M. D. (2015). Established management accounting knowledge. *Journal of Management Accounting Research*, 27(1), 123-132. <https://doi.org/10.2308/jmar-51057>
- Simons, R. (1995). Control in an age of empowerment. *Harvard Business Review*, 73(2). <http://www.hbs.edu/faculty/Pages/item.aspx?num=1904>
- Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics*, 87(3), 355-374. <https://doi.org/10.2307/1882010>
- Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.
- Venturini, L. D. B., Bianchi, M., Machado, V. N., & Paulo, E. (2022). Informational content of key audit matters and financial analysts' forecasts. *Revista Contabilidade & Finanças*, 33(89), 281-299. <https://doi.org/10.1590/1808-057x202113080>
- Wilson, D., & Sperber, D. (2004). Relevance theory. In L. R. Horn & G. Ward (Eds.), *The handbook of pragmatics* (pp. 607-632). Blackwell.
- Zafar, F., Sabri, R., Siddiqui, H. M. A., & Masood, I. (2024). Emerging issues in management accounting: Digital technologies, governance, and sustainability. *Bulletin of Business and Economics (BBE)*, 13(1). <https://doi.org/10.61506/01.00214>
- Zhang, Q., Cao, M., Zhang, F., Liu, J., & Li, X. (2020). Effects of corporate social responsibility on customer satisfaction and organizational attractiveness: A signaling perspective. *Business Ethics*, 29(1), 20-34. <https://doi.org/10.1111/beer.12243>

AUTHOR CONTRIBUTIONS

Renato Gomes de Oliveira

Conceptualization: equal;
 Data curation: lead;
 Formal analysis: equal;
 Investigation: lead;
 Methodology: lead;
 Project administration: lead;
 Software: lead;
 Supervision: equal;
 Validation: equal;
 Visualization: lead/equal/supporting;
 Writing – original draft: lead;
 Writing – review & editing: equal.

Fabírcia Silva da Rosa

Conceptualization: equal;
 Formal analysis: equal;
 Supervision: equal;
 Validation: equal;
 Visualization: equal;
 Writing – review & editing: equal.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

DATA AVAILABILITY STATEMENT

The entire dataset supporting the results of this study is available at the link:
https://github.com/renatogo-sc/pln_contabilidade_gerencia

GENERATIVE AI DISCLOSURE

The authors declare that generative artificial intelligence was used in the following stages of the production of this manuscript:

- Text refinement: Grok.
- Generation/optimization of formulas and code: Grok.
- Data analysis: Grok.

The authors declare that, regardless of the use of the tools mentioned above, all generated content was supervised, verified, and critically validated by humans. The authors assume full and exclusive responsibility for the accuracy of the data, integrity of mathematical/statistical formulas, originality of the text, and the conclusions presented in the published article.

This preprint was submitted under the following conditions:

- The authors declare that the necessary Terms of Free and Informed Consent of participants or patients in the research were obtained and are described in the manuscript, when applicable.
- The authors declare that the preparation of the manuscript followed the ethical norms of scientific communication.
- The authors declare that they are aware that they are solely responsible for the content of the preprint and that the deposit in SciELO Preprints does not mean any commitment on the part of SciELO, except its preservation and dissemination.
- The authors declare that the data, applications, and other content underlying the manuscript are referenced.
- The deposited manuscript is in PDF format.
- The authors declare that the research that originated the manuscript followed good ethical practices and that the necessary approvals from research ethics committees, when applicable, are described in the manuscript.
- The authors declare that once a manuscript is posted on the SciELO Preprints server, it can only be taken down on request to the SciELO Preprints server Editorial Secretariat, who will post a retraction notice in its place.
- The authors agree that the approved manuscript will be made available under a [Creative Commons CC-BY](#) license.
- The submitting author declares that the contributions of all authors and conflict of interest statement are included explicitly and in specific sections of the manuscript.
- The authors declare that the manuscript was not deposited and/or previously made available on another preprint server or published by a journal.
- If the manuscript is being reviewed or being prepared for publishing but not yet published by a journal, the authors declare that they have received authorization from the journal to make this deposit.
- The submitting author declares that all authors of the manuscript agree with the submission to SciELO Preprints.