Narratives of the Big-4 Transparency Reports: Country Effects or Firm Strategy?

Ana Zorio Grima
Pedro Carmona Ibáñez
Universidad de Valencia

ABSTRACT:
Purpose: The objective of this paper is to examine whether audit firms use Transparency Reports (TRs) as a tool to standardize their brand image, or whether the semantic and content analysis in these reports indicates a higher importance of country effects.

Design: Our sample includes 28 TRs published in English by the Big-4 audit firms from five EU countries (the United Kingdom, Ireland, Luxemburg, Hungary, and Malta) as well as in the US and Australia.

Findings: Using content analysis, this research finds that there is variation in the language used in TRs both across audit firms and jurisdictions. Most TRs from different countries of the same firm tend to be clustered together, suggesting that audit firms use transparency reporting as a strategy to differentiate themselves from their competitors. In fact, EY and KPMG seem to have more standardized internal procedures and standardized information. Regarding country effects, the results indicate that TRs in the UK are longer and show more detailed information.

Originality/value: Overall, this research is innovative in the sense that it applies a new methodological approach to an emerging topic such as audit transparency reporting. It identifies emerging topics of voluntary disclosure, such as financial data of the firm, gender and ethnic origin of employees, community involvement or sanctions, amongst other topics of interest which might be explored in detail by future research in order to understand the construction of the profession.

Keywords: *transparency report; audit; text mining; semantic analysis; content analysis*
Data availability: *All data are available from sources indicated in the paper.*

I. Introduction

Financial scandals have put the accounting profession under the spotlight on many occasions along the years. In fact, the last financial crisis in the United States (US) and the European Union (EU) showed once more that auditing is crucial for the reliability of financial statements and their credibility amongst investors. Subsequently, important audit reforms took place to reinforce its role (Humphrey *et al.,* 2009). As a consequence, some audit firms are now required in many jurisdictions to prepare a Transparency Report (TR) or do so on a voluntary basis. The TR should contain information about the firm's governance, structure and procedures.

The concept of transparency has gained a lot of attention in recent years, because of its presumed effects of enhancing trust and credibility (Rawlins, 2008). Nevertheless, the attempts to measure transparency in the field of audit firm reporting are limited. The narratives used in TRs as transparency practices may have no impact "if none is looking in" (Johnston, 2015) but, for those "looking in", significance of country or firm effects can be relevant for decision making and also for understanding changes in the accounting domain. In fact, processes of institutionalisation through narratives are not complete in transparency reporting. Audit firms might also disclose additional information to demonstrate an appropriate behaviour, change perceptions of their activities and social expectations about them (as discussed by Higgins *et al*., 2014 in the field of Integrated Reporting). Using content analysis, this article looks into whether the communication strategies of the Big-4 audit firms (i.e., PWC, EY, Deloitte and KPMG) make use of the benefits of this type of disclosure and identifies emerging areas of voluntary disclosure.

Audit firms' enhanced transparency should allow market participants to differentiate amongst different suppliers of audit services, providing incentives for audit firms to increase audit quality (Deumes *et al.,* 2012). This is the reason why regulators in many jurisdictions have recently introduced the requirement for some audit firms to publish a TR that contains information on audit firm governance and procedures.

However, there are reasons to suspect that the firm's communication strategy may affect the narratives used in the TR. Indeed, the firms' communication strategies may lead to standardization as the endogenous model of professional and institutional accounting expertise and technology develops, especially in the Big-4 firms. Existing research has evidenced that "professionalism" as a construct has moved from a normative-ethical approach towards an expertise approach, in large organizations or trans-national contexts (Suddaby *et al.,* 2015). In fact, accounting work can be understood as an expansion of cultural rationalization (Meyer, 1986). Therefore, micro-level dynamics generated by a standardization of the TR could imply changes in how auditing expertise and daily routines of individual professionals change. Potter (2005) highlights the importance of the processes and implications of changing the accounting domain. The TR could be a tool to expand the domain of accounting (moving away from professionalism as a normative-ethical construct towards professionalism understood as a jurisdictional claim of technical expertise – see Suddaby *et al.,* 2015), with the rationale of associating greater transparency and international comparability with more efficient capital markets.

This research sheds some light on this idea of changing the accounting domain by looking into TR. Instead of focusing attention on the extension of the accounting domain by examining new areas of professional practice (Suddaby *et al.,* 2015), this paper focus on how the publication of TR reflects a range of practices or concerns with country or firm effects that may imply a broader domain change in the profession. This is an emergent field of research as

access to these reports is rather new, and because it falls within the field of ethics, which should be preserved in all business in general, but especially in the realm of auditing as it gives credibility to the companies' financial information. Even though the TR might be seen as a transaction cost imposed on audit firms, the fact that some firms do it voluntarily or provide more information than required suggests that the firms find it useful. This analysis includes 28 TR of Big-4 firms located in seven jurisdictions: five EU countries (i.e., United Kingdom, Ireland, Luxemburg, Hungary, and Malta) as well as US and Australia. The choice of countries reflects the use of English as the language of the TR issued, because the methodological analysis employed requires the restriction to one single language, and English is the language most commonly used in the jurisdictions where TR are published.

Many accounting studies use textual analysis (for instance, Fisher *et al.,* 2010; Li, 2010; Schleicher and Walker, 2010; Boritz *et al.*, 2013; Adhikari *et al.,* 2014; Suddaby *et al.,* 2015; Boritz *et al.*, 2016; Liu, 2016; Fisher *et al.,* 2016; Zhang *et al.,* 2017) yet this study is the first to examine the semantic approach (or text mining) in TR. This paper uses content analysis, systematically transforming a large amount of text into a highly organized and concise summary of key results. It is a quantitative method for summarizing any form of content by counting various aspects of the content. The results of content analysis can be displayed with word-clouds figures. In addition, we use Latent Semantic Analysis (LSA) and hierarchical clustering- a very recent approach and therefore a significant contribution in the TR field. More precisely, this article looks into whether the similarities identified are due to regulatory frameworks (as suggested by Girdhar and Jeppesen, 2018) or the audit firm's communication strategy.

The results obtained show that the Big-4 audit firms use transparency reporting as a differentiation strategy, as the narratives of most of the TR from the same firm tend to be clustered together. Regarding country effects, the findings of this study indicate that TRs in the UK are longer and show more detailed information.

This paper has relevant policy implications in the sense that it identifies new voluntary trends of disclosure, which could bring new ideas to update the legal contents of the TR in order to promote general awareness and help create a more equal and fair society. This study is also valuable in the sense that it uses a new methodology (text mining) in the area of audit transparency reporting, and introduces in the auditing-narratives research field the evidence of standardization which could somehow reflect endogenous change (Suddaby *et al.,* 2015).

The remainder of this paper is organized as follows. Section II describes the background of this study and develops the underlying theory and research questions. Section III outlines the research design, including the description of the sample and the methodology employed. Section IV analyses the results obtained and Section V presents the conclusions and limitations of this study.

II. Background and development of research questions

The pioneer jurisdiction to require audit firms to publish a TR was the EU (Humphrey *et al.,* 2011; Maijoor and Vanstraelen, 2012). The Auditing Directive 2006/43/CE states that statutory auditors of Public Interest Entities (PIEs) have to publish a TR annually. The EU Accounting Directive 2013/34/EU updates the definition of PIEs as listed companies, credit institutions, and insurance undertakings, as well as any other undertakings designated by Member States because of their public relevance — due to their business nature, their size, or number of employees. The TR discloses certain information on the legal structure and ownership of the firm, its governance structure, internal quality control system, quality assurance, education and independence practices, as well as information on partner

remuneration (see article 40). The contents of the TR do not significantly differ across the EU Member States (Shong *et al.,* 2015), even though there are slight differences derived from the national transposition processes of the Directive. The Directive came into force as early as 2008 in some EU countries, with Ireland and Spain being the latest EU Member States to report full transposition in 2010 (EU, 2010). Each Member State has its own body that supervises the TR– e.g., the Financial Reporting Council in the United Kingdom. In fact, this body issues a publicly available report on this assessment exercise, concluding in the most recent edition that the 2013/14 TRs have significantly higher quality than those reviewed in 2010, as more firms provide key performance indicators, or describe in greater detail their processes to achieve a high audit quality (FRC, 2015).

In the European Union, new developments have taken place regarding the TR because of the 2014 audit reform. The new EU Regulation 537/2014 focuses on the specific requirements of the statutory audit of PIEs and the new Directive 2014/56/EU has amended Directive 2006/43/EC for the remaining audit firms. Most of the new provisions in the Regulation came into force in 2016. Article 13 of the new Regulation has introduced the requirement to keep the TR available online on the audit firm's website at least for five years and to communicate to the competent authorities that the TR has become available (or updated) on the website of the statutory auditor or the audit firm. In addition, it has included additional disclosures, such as a description of the statutory auditor's or the audit firm's policy concerning the rotation of key audit partners and staff, as well as information about the total revenue of the statutory auditor or the audit firm. The firm has to show the revenue split into four categories: (i) revenues from the statutory audit of annual and consolidated financial statements of PIEs and entities belonging to a group of undertakings whose parent undertaking is a PIE; (ii) revenues from the statutory audit of annual and consolidated financial statements of other entities; (iii) revenues from permitted non-audit services to entities that are audited by the statutory auditor or the audit firm; and (iv) revenues from non-audit services to other entities.

Besides the EU, the debate on audit firm's transparency has also taken place in other countries. In 2008, the US Advisory Committee on the Auditing Profession recommended that Public Company Accounting Oversight Board (PCAOB) expand transparent reporting of audit firms' governance (Bedard *et al.,* 2010; Franzel, 2016), but this type of disclosure has not been required in the form of a TR. However, the Centre for Audit Quality (CAQ) of the American Institute of Certified Public Accountants (AICPA) in 2014 recommended increased disclosures and proposed a voluntary model of TR for audit firms, which is an updated version of the model published in 2010 (AICPA, 2010; Center for Audit Quality, 2013; AICPA, 2014; Shong *et al.,* 2015). In fact, the Big-4 firms in the US voluntarily disclose an annual TR. Deloitte was the first audit firm to do so in 2010 (Bedard *et al.,* 2010), as a signal to the market of the quality of their audit services. However, the high costs of information production and proprietary disclosure involved may outweigh the benefits of a voluntary TR. As the perceived benefit is greater than the cost for audit firms with many audit clients (Fu *et al.,* 2015), the Big-4 and large second-tier audit firms issue voluntary reports worldwide.

Australia is also one of the most important jurisdictions requiring TR. Since 2013, the Australian Corporations Act 2001 s332A, as amended by the Corporations Legislation Amendment (Audit Enhancement) Act 2012, requires auditors or audit firms that audit ten or more significant entities such as listed companies or listed registered schemes, authorized deposit-taking institutions, and insurance companies to disclose an annual TR. The Australian Securities and Investments Commission reviews these TR (ASIC, 2013) but only releases the results of the inspection to the audit firm concerned because of the confidentiality provisions.

Before 2013, some Australian audit firms disclosed a TR on a voluntary basis (Oh and Dowling, 2014).

Table 1 presents a summary of the TR legal environment in the jurisdictions included in this study.

[Table 1 here]

TR can be a useful tool not only for transparency but also for commercial, marketing, or professional purposes (Oh and Dowling, 2014). Indeed, according to the accounting narratives literature (Beattie, 2014), it could also be used for opportunistic discretionary disclosure, symbolic management, reporting bias or even "cheap talk" (hyperbole that the readers easily identify and simply ignore), according to Merkl-Davies and Brennan (2011)´s terminology for impression management in a corporate reporting context.

In fact, some investors and audit committees have complained that the TR might simply include marketing information (IOSCO, 2015). In order to shed some light on this issue, semantic analysis is used to identify whether TR is used as a tool within the firm's communication strategy or whether the country's regulatory framework itself determines the similarity amongst TR. Indeed, the International Auditing and Assurance Standards Board (IAASB) states in its audit quality framework that the TR gives audit firms the opportunity to distinguish themselves by means of highlighting their own policies and audit approach, which will lead them to compete on audit quality (IAASB, 2014). On the other hand, this opportunity might make the TR a "boilerplate" document with the consequent prejudice readers might have (POB, 2010; Fu *et al.,* 2015). Apart from the minimum disclosure contents specified in some jurisdictions, the TR has a discretionary nature. Regulators do not favor the implementation of a TR template (POB, 2010), as that decision would be to the detriment of innovative formats that firms may develop for their TR (Deumes *et al.,* 2012; La Rosa *et al.,* 2018). This view is in line with the IAASB approach that TR may assist the users of audited financial statements to understand the characteristics of audit firms and their drivers for audit quality as well as assist entities in selecting a new audit firm (IAASB 2014, para 37).

Existing research on transparency reporting is rather scarce. The early study by Pott *et al.* (2008) analyzes how US practitioners assess the effectiveness of several aspects of TR in improving auditor independence and concludes that there is no significant perceived difference no matter whether the TR is mandatory or voluntary.

Deumes *et al.* (2012) examine 103 TR from four European countries (Austria, Germany, the Netherlands, and the United Kingdom) and find evidence for variation in the extent and type of disclosure across audit firms. In addition, they conclude that the quality of disclosures is not associated with audit quality, proxied by the abnormal accruals of the audit firms' clients. La Rosa *et al.* (2018) analyze 122 TR from 2010-2012 from ten European countries and find that after the release of the TR investors' confidence increases.

Some studies explicitly focus on content analysis of the TR and its extension (i.e. length in number of words or pages). Fu *et al.* (2015) present a content analysis of the 21 TR issued by Australian audit firms from 2013. They find that all the firms comply with the minimum disclosure requirements, but they show different levels of detail in some areas such as remuneration structures, the internal quality control system, or the continuing education policy, which may affect audit quality. Girdhar and Jeppesen (2018) undertake a content analysis of just 12 TR (from the Big-4 from UK, Germany and Denmark), complemented with semi-structured interviews to make cross-country comparisons. Other studies focus on Croatia (Pivac and Čular, 2012; Čular, 2014) or Spain (Zorio-Grima *et al.*, 2018), so their samples are very limited in size and their conclusions can hardly be extrapolated to other regions. Girdhar (2015)

looks into the 2012 TR of the Big-4 in the United Kingdom, Germany, and Denmark, and interviews their preparers to conclude that there are different levels of disclosure in these countries because of the transposition of the 2006 Directive. The main strength of these studies is that they are pioneers in the study of the TR from different perspectives and in different settings. However, no attention has been paid yet to TR narratives. This paper explores whether publication of TR reflects a range of practices or concerns with country or firm effects that may imply a broader domain change in the profession. Its contribution to literature is to use semantic analysis and machine learning techniques together with more traditional techniques such as Kruskal-Wallis and contingency tests to answer the following research questions:

RQ1: Are the TR language variations associated with the jurisdiction (regulatory framework explanation) or with the individual firm (communication strategy explanation)?

RQ2: Are there differentiating expressions in the TRs? Can machine-learning techniques identify the firm authoring the report by these expressions?

RQ3: Are there significant differences regarding other characteristics of the TR such as extension, presentation, or contents?

This approach renders valuable comments on emerging types of voluntary information disclosed in the TR. These new insights help to understand the profession, and give visibility to their good practices in order to construct a fairer society (Guthrie and Parker, 2017).

III. Research design

Data and Method

This analysis is based on 28 TR issued by Big-4 audit firms in five EU countries (United Kingdom, Ireland, Luxemburg, Hungary, and Malta, coded in the tables and figures below as UK, IR, LU, HU and MT, respectively) and in the US and Australia (AT), for the latest financial year available at the time of data gathering (2014/2015). Four reports are analyzed for each of these seven countries (i.e., one from each Big4). The TRs were obtained from the websites of the audit firms. As a necessary requirement for performing a consistent semantic analysis, all reports are written in English, which explains the limited number of countries in the sample.

This paper investigates whether the contents of the TRs reflect a differentiation strategy (i.e., audit firms use specific narratives that are clearly distinct from their competitors) or whether these reports indicate a higher importance of country effects (i.e., TRs are similar within one country, but there are differences across countries). Different text-mining methodologies are applied, such as bag-of-words or frequency table and latent semantic analysis.

Text-mining —a subdomain of data mining— is the analysis of natural language text (Daróczi, 2015), that is, the sequence of collecting, cleaning, processing, analyzing, and gaining useful insights from data (Aggarwal, 2015). Text-mining requires extracting meaningful numeric keys from unstructured textual information to make information manageable to the various data mining algorithms. In general, it is an interdisciplinary field with contributions from many areas, such as statistics, machine learning, information retrieval, pattern recognition, and bioinformatics. Data mining is widely used in many domains, such as retail, finance, telecommunication, and social media (Zhao, 2013). Classical applications to text-mining come from the data mining community, like document clustering and document classification. The idea is to transform the text into a structured format based on term frequencies, and subsequently apply standard data mining techniques (Witten and Frank, 2005; Feinerer *et al.,* 2008; Graham, 2016). This study focuses on understanding semantic similarities of different documents, firstly through the bag-of-words model followed by the analysis of the relationships

between the TRs in the sample based on the words or terms and latently similar words or terms they share to identify differences across audit firms (differentiation strategy) or countries. Additionally, using multidimensional scaling, a statistical technique for exploring similarities, documents are scaled and visualized on a two- and three- dimension space. All of the analyses are made in R (R Core Team, 2016) version 3.3.1, using for text-mining the tm R package version 0.6-2 (Feinerer and Hornik, 2015), for latent semantic analysis the lsa R package version 0.73.1 (Fridolin, 2015), and for stemming the SnowballC R package version 0.5.1 (Bouchet-Valat, 2014).

A bag-of-words model gives a good understanding of the content of the text and cluster analysis allows grouping documents according to their content similarities. Bag-of-words is one of the simplest language models used in natural language processing as a method of document classification (Aggarwal and Zhai, 2012). It makes a representation of the text by keeping track of the number of occurrences of each word. In this model, documents are represented as the bag (multiset) of its words (i.e., word stems). The reason to make a bag-of-words model is that it might provide a better understanding of the content of the TR and it can be used to select the features for clustering purposes.

The function *DocumentTermMatrix()* from the R package tm facilitates the building of a document-term matrix. It describes the relationship between documents and words or terms. Each row represents a document and each column a term. It displays the amount of occurrences of terms in a given document. Table 2 shows some columns and rows of the document-term matrix resulting for the sample.

Within the "bag-of-words" realm, there are several techniques for classifying common themes in documents or simply identify themes within a corpus of documents. These techniques attempt to reduce the dimensionality of the term-document matrix, such as LSA. With the LSA methodology, the term-document matrix is reduced mathematically through singular value decomposition. The merits or distinguishing feature of LSA is that it can avoid the limitations of count-based methods associated with synonyms and polysemy (terms with multiple meanings) (Loughran and McDonald, 2016).

[Table 2 here]

As can be seen in Table 2, the document-term matrix is relatively sparse or empty; there are words or terms (in the sense of word stems) that are repeated just once or even never. The next step thus is to clean the data set, that is, to remove sparse terms from the matrix (i.e., terms repeated less than once in the document). Cleansing and filtering the data included several steps, such as removing punctuation, removing numbers, converting to lowercase, removing "stopwords" (common words) and particular words and combining words that should stay together. Removing sparsity allows representing documents as vectors of significant features.

With the final matrix, word clouds, clustering techniques and machine learning techniques are applied to answer RQ1 and RQ2. The methodology used to answer RQ3 is based on content analysis.

IV. Results

Figure 1 shows in a word cloud the most frequent terms in all the TRs. The cloud gives a larger prominence to those words (or terms, in the sense of word stems) that appear more frequently. Thus, word clouds give a quick overview of frequently appearing words and allow comparisons across different text sources. As can be seen, the TRs in the sample use the word "audit" the most, followed by "firm", "quality", "partner", and "member", which seems to

highlight the importance given to the organizational structure of the firm. Other concepts of interest include "independence", "limit", "review" or "profession", which are indeed crucial for the credibility of the audit.

[Figure 1 here]

The mining process also requires designing effective analytical methods from the processed data (Aggarwal, 2015). To better reflect how important a word within a given document (TR), the term frequency–inverse document frequency (TF-IDF) is used as a weighting factor in information retrieval. The inverse document frequency is a measure of how rare or common a word is within a corpus. By comparing the TF-IDF weights to the raw term frequencies, words (or terms in the sense of word stems) that are common to many of the documents in the corpus are penalized (Hackeling, 2014). In summary, the idea of TF-IDF is to remove the effect of function words from the analysis. As a result, it reduces the weight of terms that occur very frequently in all the TRs, increasing the weight of terms that occur rarely in all TRs. Also, the TF-IDF methodology avoids the impact of the varying length of TRs, since the importance of the word increases proportionally to the number of times a word appears in the document (TF stands for Term Frequency- i.e., the number of times a word appears in that document) but is offset by the frequency of the word in the corpus (IDF is the Inverse Document Frequency, computed as the logarithm of the number of the documents in the corpus divided by the number of documents where the specific term appears). Function or common words and very frequent terms typically show up a lot in all documents, having thus a high document frequency and a low TF-IDF.

Results regarding RQ1

Using cluster analysis techniques, the document reports are grouped, putting together the semantically similar ones. Cluster analysis is a machine learning methodology that is unsupervised, that is, it does not assume a desired or ideal output corresponding to each input (James *et al.,* 2013). In this study, subgroups among all of the TRs are identified to determine different semantic territories. In other words, this study explores whether the TRs can be classified into certain categories, grouping similar documents together.

This study uses two of many techniques to semantically analyze a corpus of documents. The first, hierarchical clustering, uses the mathematical distance between the term document matrices of the TRs to assign the TRs to clusters: the greater the distance between groups of TRs, the more well-defined the clusters. Figure 2 shows the dendrogram of the corpus of reports that includes all the TRs in the sample; it results in a tree-based representation of the documents. As the results show, similarities appear among some audit firms; the groups are not clearly separated by countries. Therefore, it seems that the TRs are semantically similar for the audit firms (especially as regards EY with all its 7 TRs in the same cluster, followed by KPMG with 6 TRs).

[Figure 2 here]

The second method used in this study to group semantically similar TRs is based on LSA. This methodology has been used in the accounting domain only in very recent years. According to Loughran and McDonald (2016), LSA was first applied in business in 2006 and

one of its first applications in accounting in 2015 examined topical differences in the content of conference calls and subsequent analyst reports. Since then, a few studies use LSA to analyze topic trends in research on accounting information systems- for instance, Guan *et al.* (2016) and Hutchinson et *al.* (2018). In another study using LSA, Yang *et al.* (2018) find that most of the top topics extracted from the management discussion and analysis are not significant to differentiate bankrupt from non-bankrupt companies. In other research, word clouds are studied. For example, Mascha et al. (2018) study audit deficiencies by internal control type and firm size, finding that PCAOB inspections detect more application-level than entity-level control audit deficiencies (Mascha *et al.,* 2018). Along similar lines, this paper uses LSA and hierarchical clustering. Figure 3 illustrates a dendrogram of the TRs in the sample that has been built calculating the Euclidean distance from a latent semantic space with the R function *lsa()* from the lsa package (Fridolin, 2015). The dendrogram is similar to the one obtained directly from the weighted bag-of-words (see Figure 2), where most TRs belonging to the same firm tend to be clustered together.

[Figure 3 here]

To improve the visualization level of individual TR similarities, multidimensional scaling (MDS) is applied to the latent semantic space (Aggarwal, 2015). MDS is a form of ordination that is similar to data clustering, in the sense that various objects are ordered on the basis of multiple variables; similar objects are placed near each other in the order and dissimilar objects are separated in the order. This process maps high-dimensional vectors of variables to a lower dimensional space while attempting to preserve pairwise distances between points. MDS is particularly appropriate for the analysis of a collection of documents (Miner *et al.,* 2012). The use of methods such as MDS improves the quality of the underlying representation after performing the reduction (Deerwester *et al.,* 1990; Aggarwal, 2001).

As a result of applying MDS, the latent semantic space of all TRs is reduced to a dimensionality of 2 ($k = 2$) and 3 ($k = 3$). Figure 4 shows all documents: (a) in a two axes plot and (b) in a three axes plot. Points are colored by audit firm. The resulting representation of all TRs from the latent semantic space allows potential areas of document similarities to be identified using the variable firm as grouping factor. The firm effect is clearly identified in Figure 4.

[Figure 4 here]

Figure 5 shows a representation where points are colored by country. It is not possible to identify clusters or groups taking the country as the reference factor. As robustness tests, and in order to contemplate the regulatory auditing framework, we have performed additional analyses (not shown in tables) to identify possible content differences: (1) US vs. rest of countries, (2) Anglosphere countries vs. rest of countries, (3) EU vs. rest of countries and (4) tax havens. The results show no differences between countries who apply International Accounting and Auditing Standards versus US; neither at the country level considering Anglosphere as group factor; neither between EU and other countries. Therefore, the regulatory framework in a supra-country level does not seem to drive the TR content, either.

[Figure 5 here]

Therefore, audit firms seem to be using TRs as a differentiation strategy, which responds to RQ1. The correlogram plot of the TRs from the weighted TF-IDF space (Figure 6) confirms this differentiation strategy. A correlogram is a graph that illustrates the correlation matrix, representing each correlation coefficient with a dot color and its size according to its value. Note that the TRs of EY firms are clearly correlated, followed by those of KPMG.

[Figure 6 here]

Results regarding RQ2

To complete the text-mining of the TRs a word cloud from the weighted TF-IDF bag-of-words has been obtained for each one of the audit firms (Figure 7). TF-IDF measures how important a word is in the collection of TRs. This technique will give an answer for RQ2.

[Figure 7 here]

Even though there are many similarities, some words or acronyms appear more frequently in some TRs than in others. Bearing in mind these word clouds, the TRs were attentively read to identify patterns. Even though there is wide dispersion in length, degree of detail in the different sections, and different vocabulary used, trends can be identified, as follows. To answer RQ2, Table 3 shows the frequency means of some selected words or expressions from the TRs by audit firm. The selection method for these words was based on the weighted TF-IDF bag-of-words and expert judgement based on manual content analysis. A unigram analysis is performed using words one by one and a bigram analysis regarding two words at a time. To examine whether the mean differences between the firms are significant, Krustal-Wallis nonparametric analyses and multiple comparison using Dunn's tests are also performed. Dunn's test shows the results among multiple pairwise comparisons after a Kruskal-Wallis analysis, with a Benjamini-Hochberg correction to control the experimentwise error rate.

[Table 3 here]

The results obtained indicate that EY pays significantly more attention than the other Big-4 audit firms to describing its policies on client acceptance (numbers 1 and 31 in Table 3) and continuance (7, 32), partner rotation (10, together with KPMG), inspection (9), progress on the Global Compact's principles (33), procedures against bribery (5) and misconduct (11), as well as the audit quality review (AQR) (3) and the supervisory function carried out by the Global Professional Practice Director (PPD) (12).

KPMG also uses some standardized acronyms to describe its audit methodology, e.g., E-audit (13), which is its electronic audit workflow, the KPMG's International Audit Manual (KAM, 19), the firm's Engagement Quality Control (EQC, 14) system, the Risk Compliance Programme (RCP, 22), the Quality Performance Review (QPR, 21) as well as the Global Compliance Review (GCR, 15) focused on governance and risk (23) amongst other processes. Regarding independence issues, KPMG has a whistle-blowing hotline (17, as well as EY) and a web-based independence tracking system (KICS, 28). KPMG is the Big-4 that emphasizes industry (25) specialization (24) the most. As could be expected, the word "Sentinel" (29) appears significantly more often in TRs by KPMG, as KPMG has this global conflicts and independence checking system in its offices worldwide.

KPMG refers significantly more often to its own Professional Judgement (26) Framework (27) and to the expression "tone at the top" (i.e., an ethical climate established by

senior management and corporate governance structures, 34) than the other firms. Note that this expression became popular following some major accounting scandals such as Enron and WorldCom, when the Sarbanes–Oxley Act of 2002 emphasized this expression as an effective way to prevent and detect financial fraud and unethical practices.

PWC refers significantly more often to some standardized acronyms than the other firms. CES (6) is a Central Entity Service with information about clients, to determine independence status before entering into a new non-audit engagement. AURA (4) refers to the PWC global Enterprise Resource Planning (ERP) system and QMR (20) to the globally standardized Quality Management Review.

Deloitte uses significantly more often (at the 10% level) the expressions GIMS (Global Independence Monitoring System) (16) and DESC (Deloitte Entity Search and Compliance System) (8), which are tools to control compliance with independence policies. Partners search in a database of internationally restricted entities (i.e., DESC) and/or in another database of financial interests and brokerage accounts (GIMS) to identify whether an entity or its financial interests or brokerage accounts pose an independence threat.

Most of the Big-4 refer to their own Code of Conduct, but KPMG and EY refer significantly more often to compliance with the Code of Ethics for Professional Accountants issued by the International Ethics Standards Board for Accountants (IESBA, 18). They also standardize their TRs and their procedures worldwide, as shown by the fact that the word "global" (30) appears significantly more often in the TRs published by these firms because of a common communication strategy regarding transparency.

Additionally, a classification model has been fitted to determine which terms and expressions included in Table 3 are relevant in relating the respective TR as belonging to a specific audit firm. Extreme Gradient Boosting (XGBoost) is the machine learning (ML) algorithm selected to learn from words or terms contained in all of the TRs. One important advantage of ML techniques is it does not require prior assumptions about the underlying relationships between the variables. XGBoost is a relatively new algorithm, but a very efficient implementation of the gradient boosting framework that can be applied in classification scenarios (Chen and Benesty, 2016). Boosting is particularly useful and frequently used in the context of classification problems, minimizing overall errors by introducing supplementary models based on the errors from previous iterations. Boosting methods ensemble several single models into a combined model, which increases the predictive power over those of the individual models (Chambers and Dinsmore, 2015). XGBoost computes the probability that a given TR belongs to each of the audit firms and assigns the TR to the firm with the highest score.

Figure 8 represents the relatively most important terms and expressions (features) for the fitted XGBoost model. The plot shows the gain contribution of each term/expression to the model. For boosted tree models, each gain of each variable of each tree is taken into account, then average per variable to give a vision of the entire model. Highest value means important variable to classify the response variable (Chen and Benesty, 2016). Here, the most important terms or expressions are consistent with those having significant mean differences between the firms (see Table 3).

[Figure 8 here]

Accuracy is a measure of model performance. Table 4 displays the number of correct classifications produced by the fitted XGBoost model as a confusion matrix or table of the predicted results. This table compares the actual outcomes to the predictions, which allows to

visualize the performance of the XGBoost technique used to fit the model. The set of probabilities obtained by the XGBoost model is transformed into a label prediction. The global percentage of accuracy or the actual cases properly classified is 100%; that is, the model classifies correctly all of the analyzed TRs as belonging to a given audit firm, which confirms the results regarding RQ2. Note that this is an exploratory analysis, so the objective is not to determine if a given model is generalizable. In other words, prediction is not the main objective of this paper, but the fact that the XGBoost model has such a high percentage of accuracy confirms our findings above on a firm effect.

[Table 4 here]

As a sensitivity analysis, the Extreme Gradient Boosting algorithm has been re-run dropping terms that are firm-specific. Nonetheless, classification accuracy is still 100% (as in Table 4), while the terms and their importance obviously change from the ones in Figure 8 as the latter considers firm-specific expressions.

Results regarding RQ3

In addition to the identification of semantic analysis differences between the TRs, this paper also analyzes other characteristics such as extension, presentation, or content analysis in accordance with RQ3. Table 5 shows the descriptive statistics based on extension and manual content analysis.

[Table 5 here]

Tables 6 and 7 show the Anova —including pairwise comparisons— and contingency tests for the continuous and dummy variables respectively of characteristics reported in Table 5. The results indicate the existence of significant differences between certain firms or countries for specific variables.

[Table 6 here]

For instance, as can be seen in Table 6, extension is significantly higher in the TRs from the UK than in those from other countries (both in *pages* and in the number of *words*), as well as in the number of paragraphs explaining the legal structure of the firm (*legal_par*). Notwithstanding that, regarding the number of paragraphs on quality control (*QC_par*), the difference is statistically significant only for KPMG (the highest number of paragraphs) and Deloitte (the lowest). With respect to the number of PIEs reported (*num_PIE*), only statistically significant differences are found regarding country comparison (being Australia, UK, and Ireland the countries where audit firms report higher numbers). No significant differences are found regarding the paragraphs devoted to the firm's policy on education (*Edu_par*), whereas in the UK, there is significantly more disclosure on the partner remuneration policy (*pay_par*). Hungary is the country where TRs report a significantly higher ratio of audit fees to total fees (*audit_total*). KPMG devotes significantly more paragraphs to the audit quality policy than EY

(*quality_par*), and both firms dedicate significantly more paragraphs to the acceptance engagement policy than PWC (*engage_par*). Finally, note that the number of letters presenting the TR is not significant but in KPMG is even higher than two (*letter_num*).

[Table 7 here]

The contingency tests reported in Table 7 show significant country differences regarding the disclosure of key financial data (*key_fin_data*). In the UK, all the Big-4 report key financial data, three in Australia, two in Malta, one in the US, and none in the other three countries. Disclosure on the exact number of years in their rotation policy is significantly more common in EY's TRs than in KPMG's (*yrs_rot*). EY is also significantly more transparent disclosing the number of minimum hours of education (in all of their TRs) than the other firms (*Edu_hours*). Disclosure of the type of revenues earned (e.g., assurance, advisory, tax) by the type of client (audit versus non-audit clients) is significantly different between the countries in the sample (*client_dist*). In the UK, all of the Big-4 provide that split and three in Australia do so, whereas reporting this information is not common in the other countries — only two TRs in Malta and one in Ireland contain this information. Even though providing comparative figures of the previous year is the practice always applied in some countries (Australia, Ireland, Malta, and the UK), there are significant differences with other countries (*comparat*). KPMG is the firm that acknowledges significantly more often that there is no fee dependence from any client (*fee_dep*). For instance, in the TR of KPMG Malta it is stated that "*No audit client accounted for more than 10 percent of the total fees received by our firm over the last two years*". Although explicitly acknowledging that "*audit partners or members of the audit team cannot be compensated for the sale of non-audit services to audit clients*" is common practice in some countries (Australia, Ireland, Luxembourg, and Malta), there are significant differences to the other countries (*more_client).* In fact, fee dependency and audit partner compensation for NAS are covered in the IFAC Code of Ethics (IESBA, 2016, para 290.215 and 290.224). There are significant differences regarding corporate social responsibility disclosures; EY is the only firm reporting its commitment to the Global Compact principles (*Global_comp*). At the country level analysis, there are differences in *gender* disclosures and *life-work* balance information (Malta is the country with more TRs disclosing this type of information). Significant differences are found for ethics because EY and PWC always refer to the code of the International Ethics Standards Board for Accountants (*iesba_code*), to the rules of the US Securities and Exchange Commission (*Sec*, always reported in the TRs of EY and KPMG), their own list of *values* (EY), or the US Public Company Accounting Oversight Board (*Pcaob*, always in the US, the UK, Luxembourg, and Ireland). It is noteworthy that EY and PWC always refer to the International Standard on Quality Control 1 (*Isqc*, whereas Deloitte never does. At the country level, only in the UK three TRs disclose information on sanctions received (*Sanctio*), or the name of the external auditor of the audit firm (*ext_audit*).

As can be seen from table 7, some variables are not significantly different between the firms or countries. Nevertheless, there is a trend to show pictures of some of the partners (*photo*), to disclose the exact number of *partners*, and to disclose the last assurance review (*QC-date*, as required in the EU 2006 Directive), and the spread use of the expression "*tone at the top*". Emerging initiatives include reporting on the *ethnic* origin of employees. This type of initiatives can be most valuable to provide data for a hot topic research such as ethnicity and gender – e.g., Huang *et al.* (2016); Hayes and Jacobs (2017). Some TRs include voluntary disclosures on community involvement (*communi*), environmental issues (*Env*),

communication with stakeholders (*Stake*), awards received, a table on partners' meeting attendance (*att_meet*), or key performance indicators (*KPI*s).

As robustness tests, and in order to contemplate the regulatory auditing framework, we have performed additional tests to identify differences regarding other characteristics of the TR such as extension, presentation, or contents: (1) US vs. rest of countries, (2) Anglosphere countries vs. rest of countries and (3) EU vs. rest of countries, (4) tax havens vs. other countries. In general, the results showed no relevant differences.

V. Conclusion

Enhanced transparency of audit firms should intensify competition in the audit market, ultimately resulting in an improvement of audit quality. However, this document may simply increase stakeholder's information overload and introduce more compliance requirements for audit firms.

This study examines 28 TRs from the Big-4 audit firms from five EU countries (United Kingdom, Ireland, Luxemburg, Hungary, and Malta) as well as the US and Australia. With the publication of the 2006 Directive, Europe has pioneered requiring audit firms that audit PIEs to issue a TR. This initiative was followed by Australia in 2012, being a controversial topic especially in the US, where there is a voluntary model for TR developed by the Center for Audit Quality (AICPA, 2010 and 2014). This paper investigates whether the language used in the TRs reveals a differentiation among audit firms or responds to jurisdictional regulatory differences. The results obtained show that there is variation in the language used in the TRs both across audit firms and across jurisdictions. However, using text-mining methodology, most TRs belonging to a same firm tend to be clustered together, suggesting that audit firms use transparency reporting in their communication strategy. Invoking Occam's Razor, one explanation may be simply a case of efficiency. It is efficient and less costly for audit firms to benefit from centralisation and use the same template for TRs. In this sense, our analysis evidences that EY and KPMG highly standardize their internal procedures and the information produced in the TR, as compared to the other Big-4.

The firm-effects regarding the TR narratives may imply the endogenous and organic adoption of key protocols within the firm, which is in line with Suddaby *et al.* (2015), who analyze the use of social media by the Big-4. In light of the findings of this research, these protocols aim for instance, at preserving independence, improving the relationship with stakeholders by disclosing information on corporate social responsibility issues, or being more transparent in reporting key financial figures or comparative information from previous years. In fact, these large multinational firms seem to be leading the "site of professional emergence" (Suddaby *et al.,* 2015). Standardization (institutional mimicry) and (commercial / economic) utilitarianism are at the basis of firms' organizational culture.

Using content analysis, significant differences are found regarding TR disclosures. At the firm level, EY is the firm disclosing significantly more specific information on its engagement policy, rotation policy, hours of education, commitment with the Global Compact, compliance with the IESBA Code of Ethics, the firm's list of values, or ISQC1. KPMG is also providing significantly more information about quality control, audit quality and engagement policy, fee-dependence, and reference to the US SEC. At the country level, TRs from the UK are significantly longer with significantly more information about the legal structure. Audit firms from the UK report that they audit a large number of PIEs, provide more details on the partner remuneration policy, more information on key financial data, the types of revenues per type of client, the name of their own external auditor and sanctions received. The latter might constitute interesting input for future research in order to explore the effects of "visibility of

sanctions" (Lesage *et al*., 2016). In Australia, all of the Big-4 report that partners do not receive extra remuneration for selling non-audit services to audit clients and the number of PIEs is the highest, followed by TRs from the UK and Ireland. Our study does not suggest special differentiated strategies regarding TR disclosures by the Big-4 in tax havens (such as Malta and Luxembourg). TR from Malta have been highlighted in several topics above (e.g., gender disclosures and life-work balance, audit partner compensation for non-audit services, fee dependence, comparative figures, client distribution, financial data), whereas that is not the case for Luxembourg. Therefore, no general implications regarding transparency by auditors in tax havens can be drawn from this research.

Likewise, our study does not suggest special differentiated strategies regarding TR disclosures by the Big-4, following the categories mentioned above (1) US vs. rest of countries, (2) Anglosphere countries vs. rest of countries and (3) EU vs. rest of countries. Since most of the TR information content is voluntary or, even if it fulfills a specific disclosure at supranational level (EU) it is under the supervision of a national regulator, firms are flexible to develop the contents further and the trends herein identified indicate country effects. Therefore, according to our results the driver for disclosures seems more related to the country business culture and national social attitudes regarding transparency.

This study has the limitation of a rather limited sample size, which is due to the need to use just one language because of the methodology employed. A further limitation is the fact that audit firms have to be present in all of the countries analyzed. Note that this limitation is inherent to the topic under study -see, for instance, Fu *et al.* (2015) with a sample of only 21 TR.

The implications of this article are interesting to both practitioners and academics alike. From a practical standpoint, our study is extremely innovative as it shows the wide range of topics dealt with in TR and gives an overview of country trends, which can be useful for future studies. No policy recommendation can be made based on our findings with respect to recommending a specific format of TR amongst the ones included in our sample, even though our impression after analyzing all these reports supports the recommendations by some regulators to disclose key performance indicators in TR to allow easier comparisons through time and audit firms (ASIC, 2013; FRC, 2015; PCAOB, 2015; CPAB, 2017).

This article is unique in the sense that it uses a new methodological approach (i.e., semantic analysis and machine learning techniques) in the area of accounting and auditing to reflect on transparency reporting by audit firms. Text mining turns text data into high quality information that is more easily analyzed that the raw text data. Consequently, it helps to minimize human effort in dealing with text data. Text mining also involves extracting relevant information from a large amount of text data, thereby turning big text data into a relatively small amount of more relevant text data. Therefore, text mining presents an opportunity for researches to make sense of vast information and data resources. This study investigated the research questions using both traditional techniques (e.g., frequency rankings, hierarchical clustering, Kruskal-Wallis contingency tests, and manual content analysis) and more recent machine learning algorithms (e.g. LSA and XGBoost). While interpretation of results when complex algorithms are used is a challenge, visualizations that reveal patterns in the data and highlight important features are provided to help with interpretation of results.

New avenues for future research may use the new information disclosures required by the EU Regulation coming into force in 2016 or focus on the information provided by the TR inspection reports by public oversight bodies. Even more importantly, future studies might also analyze if TR are really used by the stakeholders and how. Suggestions have been made that these reports include commercial or marketing-like materials (Oh and Dowling, 2014; IOSCO,

2015, CPAB, 2016). If so, this might lead to the idea that TR are just a bureaucratic imposition from regulators who need to be seen to be responding to the Global Financial Crisis, and that TR are generally not widely read or taken terribly seriously. As regards the former, our findings limited to the Big-4 suggests that the cost-benefit analysis of publishing the TR is positive, as they do so even in a voluntary basis – e.g., in the US. As regards the latter, we think that there is a wide range of potential users of this report who can find it useful, apart from regulators. Even though literature often refers to isomorphism in auditor's choice (Corten 2018), audit committees and management are the user groups who are expected to find TR more relevant in order to make the initial selection decision and subsequent retention evaluation process. Investors are expected to benefit from TR because these reports are supposed to promote high-quality audits. Prospective employees of an audit firm may find some pieces of information contained in the TR very interesting- e.g. the information regarding the hours of education, gender and racial equality policies, life-working balance or salaries and remuneration policies and can be useful to prevent future decisions to leave the audit profession (Gertsson *et al.*, 2017). The management of the audit firms is also a very relevant user of TR, as this report may make them reflect on the firm's current policies on the many issues impacting audit quality and how to improve them in the future as well as compare itself with the competition. In this sense, it is relevant to highlight that some regulators are checking not only the mere existence of TR but also revising its contents (e.g., EU Regulation 537/2014, art 26). In Australia, the regulator also reviews selected TR, seeking further explanations if there are inconsistencies with the knowledge and experience obtained from its inspections of audit firms (ASIC, 2013). Moreover, the Financial Review Council, which is in charge of the review of the TRs in the UK, suggests that firms should consider providing details in their TR on their KPIs and progress on KPIs in future TRs (FRC 2015). Regulators in Canada, Australia and the US are also discussing or promoting the disclosure of Audit Quality Indicators (AQI) (ASIC, 2013; PCAOB, 2015; CPAB, 2017). Therefore, academic research will soon be able to take advantage of new data to help understand the construction of the profession and identify good practices. All these new developments are aimed at increasing investor's trust in audited financial reports and suggest that auditors' transparency reports are likely to be in the spotlight in the coming years.

References

Adhikari, A., Betancourt, L., and Alshameri, F. (2014). The SEC's proposed IFRS roadmap: An analysis of comment letters using content analysis and textual software. *Journal of International Accounting, Auditing and Taxation*, Vol. 23, No.2, pp. 98-108.

Aggarwal, C. C. (2001). On the effects of dimensionality reduction on high dimensional similarity search. *ACM PODS Conference*, 256–266.

Aggarwal, C. C. (2015). *Data Mining: The Textbook*. Springer International Publishing. Switzerland.

Aggarwal, C. C., and Zhai, C. (2012). *Mining text data*. Springer Science and Business Media.

AICPA, American Institute of Certified Public Accountants (2010). *Voluntary Model U.S. Transparency Report.* Center for Audit Quality of AICPA. Published on April 06, 2010.

AICPA, American Institute of Certified Public Accountants (2014). *Voluntary Model U.S. Transparency Report.* Center for Audit Quality of AICPA. Published on July 23, 2014.

ASIC, Australia Securities and Investment Commission (2013). *Information sheet 184: Audit transparency reports*. Published in August 2013. Available at: http://download.asic.gov.au/media/1338902/info184-published-19-August-2013.pdf reports/ (accessed 2 February 2018).

Beattie, V. (2014). Accounting narratives and the narrative turn in accounting research: Issues, theory, methodology, methods and a research framework. *The British Accounting Review*, Vol. 46, No.2, pp. 111–134.

Bedard, J., Johnstone, K. and Smith, E. (2010). Audit Quality Indicators: A Status Update on Possible Public Disclosures and Insights from Audit Practice. *Current Issues in Auditing*. American Accounting Association, Vol. 4, No.1, C12-C19.

Bouchet-Valat, M. (2014). *SnowballC: Snowball stemmers based on the C libstemmer UTF-8 library*. R package version 0.5.1. https://CRAN.R-project.org/package=SnowballC.

Boritz, J.E., Hayes L. and Lim, J. (2013) A Content Analysis of Auditors' Reports on IT Internal Control Weaknesses: The comparative advantages of an automated approach to control weakness identification. *International Journal of Accounting Information Systems* Vol. 14, No. 2, pp. 138-163.

Boritz, J.E., Hayes, L. and Timoshenko, L. (2016). Determinants of the Readability of SOX 404 Reports. *Journal of Emerging Technologies in Accounting* Vol. 13, No. 2, pp. 145-168.

Center for Audit Quality (2013). *CAQ Resource: Audit Quality Reporting*, Aug. 20, 2013. Available at : http://thecaq.org/file/1316/download?token=ObEE6oH4 (accessed 30 March 2016).

Chambers, M. and Dinsmore, T. W. (2015). *Advanced Analytics Methodologies: Driving Business Value with Analytics*. Pearson Education. New Jersey.

Chen, T. and Benesty, M. (2016). *xgboost: eXtreme Gradient Boosting.* R package version 0.4-3. https://cran.r-project.org/web/packages/xgboost/vignettes/xgboost.pdf

Corten, M., Steijvers, T., & Lybaert, N. (2018). Auditor choice in private firms: a stakeholders perspective.*Managerial Auditing Journal*, Vol. 33, No 2, pp. 146-170.

CPAB, Canadian Public Accountability Board. (2016). *Transparency into audit- Audit Quality Indicators and Transparency Reporting*. March 2016. Available at : http://www.cpab-ccrc.ca/Documents/Stakeholders/Audit%20Committee/CPAB_Exchange_Transparency_Audit_Part_2_EN.pdf (accessed 30 January 2018).

CPAB, Canadian Public Accountability Board (2017). *2016 Audit Quality Indicators Pilot Project Interim Report Summary*. Available at: http://www.cpab-ccrc.ca/Documents/Topics/Audit%20Quality%20Indicators/CPAB_Exchange_AQI_Pilot_Summary_EN.pdf (accessed 30 January 2018).

Čular, M. (2014). Transparency of audit firms in Croatia. *Special Journal Issue on Advances in Business, Finance, Economics and Managemen*t, Vol. 8, No.3, pp. 324-328.

Daróczi, G. (2015). *Mastering Data Analysis with R*. Packt Publishing. Birmingham, UK.

Deerwester, S., Dumais, S.T., Landauer, T., Furnas G., and Harshman, R. (1990). Indexing by latent semantic analysis. *JASIS*, Vol. 41, No.6, pp. 391–407.

Deumes, R., Schelleman, C., Vander Bauwhede, H. and Vanstraelen, A. (2012). Audit firm governance: Do transparency reports Reveal Audit Quality?. *Auditing: A Journal of Practice & Theory*, Vol. 31, No.4, pp., 193-214.

EU, European Union (2010). *Scoreboard on the transposition of the Statutory Audit Directive (2006/43/EC). 1 September* 2010. Available at http://ec.europa.eu/internal_market/auditing/docs/dir/01_09_10_scoreboard_en.pdf (accessed 30 March 2016).

Feinerer, I. and Hornik, K. (2015). *tm: Text Mining Package. R package version 0.6-2.* https://cran.r-project.org/package=tm

Feinerer, I., Hornik, K. and Meyer, D. (2008). Text Mining Infrastructure in R. *Journal of Statistical Software*, Vol. 25, No.5, pp. 1–54.

Fisher, I. E., Garnsey, M. R., Goel, S., and Tam, K. (2010). The role of text analytics and information retrieval in the accounting domain. *Journal of emerging technologies in Accounting*,Vol. 7, No. 1, pp. 1-24.

Fisher, I. E., Garnsey, M. R., and Hughes, M. E. (2016). Natural language processing in accounting, auditing and finance: A synthesis of the literature with a roadmap for future research. *Intelligent Systems in Accounting, Finance and Management*, Vol. 23, No. 3, pp. 157-214.

Franzel, J. M. (2016). Are we there yet? Protecting investors by securing a strong auditing profession into the future. *Research in Accounting Regulation*, Vol. 28, No.1, pp.42-54.

FRC, Financial Reporting Council (2015). *Transparency Reporting by Auditors of Public Interest Entities Review of Mandatory Reports*. March 2015. Available at : https://www.frc.org.uk/getattachment/48ba388f-4f6b-40d9-9323-e1683f1d0732/Transparency-reporting-review-of-mandatory-reports-2015.pdf (accessed 2 February 2018).

Fridolin, W. (2015). *lsa: Latent Semantic Analysis. R package version 0.73.1*. https://cran.r-project.org/package=lsa

Fu, Y., Carson, E., and Simnett, R. (2015). Transparency report disclosure by Australian audit firms and opportunities for research. *Managerial Auditing Journal*, Vol. 30, No 8/9, pp.870-910.

Gertsson, N., Sylvander, J., Broberg, P., and Friberg, J. (2017). Exploring audit assistants' decision to leave the audit profession. *Managerial Auditing Journal*, Vol. *32*, No 9, pp. 879-898.

Girdhar, S. (2015). *The Internationalization of Big Accounting Firms and the Implications on their Practices and Structures: An Institutional Analysis*. Doctoral dissertation, Aarhus UniversitetAarhus University, Aarhus BSSSchool of Business and Social Sciences, Institut for ØkonomiDepartment of Economics and Business Economics.

Girdhar, S., and Jeppesen, K. K. (2018). Practice variation in Big-4 transparency reports. Accounting, Auditing & Accountability Journal, Vol. 31, No 1, pp. 261-285.

Graham, W. (2016). *Hands-On Data Science with R Text Mining*. Retrieved January 28, 2017, from http://handsondatascience.com/TextMiningO.pdf

Guan, J., Levitan, A. S., and Goyal, S. (2016). Text mining using latent semantic analysis: An illustration through examination of 30 years of research at JIS. *Journal of Information Systems*, Vol 32, No. 1, pp.67-86.

Guthrie, J., and Parker, L. D. (2017). Reflections and projections: 30 years of the interdisciplinary accounting, auditing and accountability search for a fairer society. *Accounting, Auditing & Accountability Journal*, Vol 30, No. 1, pp.2-17.

Hackeling, G. (2014). *Mastering Machine Learning with scikit-learn*. Packt Publishing Ltd.

Hayes, C. and Jacobs, K. (2017). The processes of inclusion and exclusion: The role of ethnicity and class in women's relation with the accounting profession. *Accounting, Auditing & Accountability Journal*, Vol. 30, No. 3, pp. 565-592.

Higgins, C., Stubbs, W., and Love, T. (2014). Walking the talk (s): Organisational narratives of integrated reporting. *Accounting, Auditing & Accountability Journal*, Vol. 27, No.7, pp. 1090-1119.

Huang, G., Fowler, C. J., and Baskerville, R. F. (2016). Entering the accounting profession: the operationalization of ethnicity-based discrimination. *Accounting, Auditing & Accountability Journal*, Vol.29, No. 8, pp. 1342-1366.

Humphrey, C., Loft, A., and Woods, M. (2009). The global audit profession and international financial architecture: Understanding regulatory relationships at a time of financial crisis. *Accounting, Organizations and Society*, Vol. 34, pp. 810-825.

Humphrey, C., Kausar, A., Loft, A., and Woods, M. (2011). Regulating Audit beyond the Crisis: A Critical Discussion of the EU Green Paper. *European Accounting Review*, Vol. 20, No.3, pp. 431-457.

Hutchison, P. D., Daigle, R. J., and George, B. (2018). Application of latent semantic analysis in AIS academic research. *International Journal of Accounting Information Systems*. In press https://doi.org/10.1016/j.accinf.2018.09.003

IAASB, International Auditing and Assurance Standards Board. (2014). "A framework for audit quality". Available at: www.ifac.org/sites/default/files/publications/files/A- Framework-for-Audit-Quality-Key-Elements-that-Create-an-Environment-for-Audit- Quality-2.pdf (accessed 25 January 2015).

IESBA, International Ethics Standards Board for Accountants (2016). Handbook of the Code od Ethics for Professional Accountants. 2016 Edition. Available at: https://www.ethicsboard.org/system/files/publications/files/2016-IESBA-Handbook.pdf (accessed 31 January 2018).

IOSCO, International Organization of Securities Commissions (2015). *Transparency of Firms that Audit Public Companies*. November, 2015. Available at: http://www.iosco.org/library/pubdocs/pdf/IOSCOPD511.pdf (accessed 25 March 2016).

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical Learning*. 2nd Edition. Springer. New York.

Johnston, M. (2015). Making transparency real? Accounting and popular participation in corruption control. *Critical Perspectives on Accounting*, Vol. 28, pp. 97-101.

La Rosa, F., Caserio, C., and Bernini, F. (2019) Corporate governance of audit firms: Assessing the usefulness of transparency reports in a Europe-wide analysis. *Corporate Governance: An International Review*. Vol. 27, pp. 14-32.

Lesage, C., Hottegindre, G., & Baker, C. R. (2016). Disciplinary practices in the French auditing profession: Serving the public interest or the private interests of the profession?. *Accounting, Auditing & Accountability Journal*, Vol. 29, No. 1, pp. 11-42.

Li, F. (2010). Textual analysis of corporate disclosures: a survey of the literature. *Journal of Accounting Literature*, Vol. 29, pp. 143–165.

Liu, Q (2016). Textual Analysis: A Burgeoning Research Area in Accounting. *Journal of Emerging Technologies in Accounting,* Vol. 13, No.2, pp. 89-91.

Loughran, T., and McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, Vol. 54, No.4, pp. 1187-1230.

Maijoor, S. J. and Vanstraelen, A. (2012). Research Opportunities in Auditing in the EU. *Auditing: A Journal of Practice & Theory,* Vol. 31, No.1, pp. 115-126.

Mascha, M. F., Lamboy-Ruiz, M. A., & Janvrin, D. J. (2018). PCAOB inspections: An analysis of entity-level and application-level control audit deficiencies. *International Journal of Accounting Information Systems*, Vol 30, pp. 19-39.

Merkl-Davies, D.M., and Brennan, N.M. (2011). A conceptual framework of impression management: new insights from psychology, sociology and critical perspectives. *Accounting and Business Research*, Vol. 41, No.5, pp. 415–437.

Meyer, J. (1986). Social environments and organizational accounting. *Accounting Organizations and Society*, Vol. 11, pp. 345–356.

Miner, G., Elder IV, J., Fast, A., Hill, T., Nisbet, R., and Delen, D. (2012). *Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications*. Academic Press.

Oh, K. and Dowling, C. (2014). "Professionalism versus commercialism: an analysis of the image conveyed by audit firms in transparency reports", Working Paper, University of Melbourne.

Pivac, S., and Čular, M. (2012). Quality Index Creating and Analysis of the Transparency of Audit Firms – Case Study in Croatia, *Croatian Operational Research Review,* Vol. 3, pp. 224 – 235.

PCAOB, Public Company Accounting Oversight Board (2015). *Concept Release on Audit Quality Indicators.* Available at: https://pcaobus.org/Rulemaking/Docket%20041/Release_2015_005.pdf (accessed 30 January 2018).

POB, Professional Oversight Board. (2010). Transparency reporting by auditors of public interest entities: review of mandatory reports. Available at: http://www.firc. org.uk/pob/publications/index. cfnUmode'/dist&cID'M. (accessed 30 March 2016).

Pott, C., Mock, J.T. and Watrin, C. (2008). *The Effect of a Transparency Report On Auditor Independence*: *Practitioners' Self-assessment*. *Review Managerial Science,* Vol. 2, No.2, pp. 111-127.

Potter, B. N. (2005). Accounting as a social and institutional practice: Perspectives to enrich our understanding of accounting change. *Abacus*, Vol. 41, pp. 265–289.

Rawlins, B. (2008). Give the emperor a mirror: Toward developing a stakeholder measurement of organizational transparency. *Journal of Public Relations Research*, Vol. 21, No.1, pp. 71-99.

R Core Team (2016). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*, Vienna, Austria. URL http://www.R-project.org/.

Schleicher, T. and Walker, M. (2010). Bias in the tone of forward-looking narratives. *Accounting and Business Research*, Vol. 40, No.4, pp. 371–390.

Shong, K. R., Yap, W. J., and Jeong, J. (2015). *ANCAAR Audit Quality Project. Phase 2 – Auditor transparency reports.* Available at: https://www.rsa.anu.edu.au/media/774108/Final-Report_Shong-Yap-Jeong.pdf (accessed on the 23 March 2016).

Suddaby, R., Saxton, G. D., and Gunz, S. (2015). Twittering change: The institutional work of domain change in accounting expertise. *Accounting, Organizations and Society*, Vol. 45, pp. 52-68.

Witten, I. and Frank E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques, second ed*. Morgan Kaufmann, San Francisco, CA, US.

Yang, F., Dolar, B., and Mo, L. (2018). Textual Analysis of Corporate Annual Disclosures: A Comparison Between Bankrupt and Non-Bankrupt Companies. *Journal of Emerging Technologies in Accounting*. Vol. 15, No. 1, pp. 45-55.

Zhao, Y. (2013). *R and Data Mining. Examples and Case Studies*. Elsevier. London, UK.

Zhang, M. C., Stone, D., & Xie, H. (2017). Text Data Sources in Archival Accounting Research: Insights and Strategies for Accounting Systems' Scholars. *Journal of Information Systems*. https://doi.org/10.2308/isys-51979

Zorio-Grima, A., García-Benau, M. A., Grau-Grau, A. J., and Paredes-Ojeda, F. (2018). El informe de transparencia de las firmas auditoras: Evidencia del mercado español 2010-2013. *Spanish Journal of Finance and Accounting/Revista Española de Financiación y Contabilidad*, Vol. 47, No 2, pp. 280-305.

TABLE 1

TR regulatory environment

JURISDICTION	NAME OF LEGAL DISPOSITION	TR PUBLICATION	TR CONTENTS
EUROPE	Introduced by the Statutory Audit Directive (2006/43/EC), transposed to national law by member countries yet superseded as regards TR, by EU Regulation 537/2014 on specific requirements regarding statutory audit of public-interest entities (article 13). TR are compulsory for statutory auditors and audit firms that carry out	TR should be published at the latest four months after the end of each financial year. That transparency report shall be published on the website of the statutory auditor or the audit firm and shall remain available on that website for at least five years from the day of its publication on the website. If the statutory auditor is employed by an audit firm, the obligations under this Article shall be incumbent on the audit firm. A statutory auditor or an audit firm shall be allowed to update its published annual transparency report. In such a case, the statutory auditor or the audit firm shall indicate that it is an updated version of the report and the original version of the report shall	The TR shall include at least the following: (a) a description of the legal structure and ownership of the audit firm; (b) where the statutory auditor or the audit firm is a member of a network: (i) a description of the network and the legal and structural arrangements in the network; (ii) the name of each statutory auditor operating as a sole practitioner or audit firm that is a member of the network; (iii) the countries in which each statutory auditor operating as a sole practitioner or audit firm that is a member of the network is qualified as a statutory auditor or has his, her or its registered office, central administration or principal place of business; (iv) the total turnover achieved by the statutory auditors operating as sole practitioners and audit firms that are members of the network, resulting from the statutory audit of annual and consolidated financial statements; (c) a description of the governance structure of the audit firm; (d) a description of the internal quality control system of the statutory auditor or of the audit firm and a statement by the administrative or management body on the effectiveness of its functioning; (e) an indication of when the last quality assurance review referred to in Article 26 was carried out; (f) a list of public-interest entities for which the statutory auditor or the audit firm carried out statutory audits during the preceding financial year;

	statutory audits of public-interest entities	continue to remain available on the website. Statutory auditors and audit firms shall communicate to the competent authorities that the transparency report has been published on the website of the statutory auditor or the audit firm or, as appropriate, that it has been updated.	(g) a statement concerning the statutory auditor's or the audit firm's independence practices which also confirms that an internal review of independence compliance has been conducted; (h) a statement on the policy followed by the statutory auditor or the audit firm concerning the continuing education of statutory auditors referred to in Article 13 of Directive 2006/43/EC; (i) information concerning the basis for the partners' remuneration in audit firms; (j) a description of the statutory auditor's or the audit firm's policy concerning the rotation of key audit partners and staff in accordance with Article 17(7); (k) where not disclosed in its financial statements within the meaning of Article 4(2) of Directive 2013/34/EU, information about the total turnover of the statutory auditor or the audit firm, divided into the following categories: (i) revenues from the statutory audit of annual and consolidated financial statements of public-interest entities and entities belonging to a group of undertakings whose parent undertaking is a public-interest entity; (ii) revenues from the statutory audit of annual and consolidated financial statements of other entities; (iii) revenues from permitted non-audit services to entities that are audited by the statutory auditor or the audit firm; and (iv) revenues from non-audit services to other entities. The statutory auditor or the audit firm may, in exceptional circumstances, decide not to disclose the information required in point (f) of the first subparagraph to the extent necessary to mitigate an imminent and significant threat to the personal security of any person. The statutory auditor or the audit firm shall be able to demonstrate to the competent authority the existence of such threat.
AUSTRALIA	Under section 322 of the Corporations Act, all individual auditors, audit firms and authorised audit companies (collectively, auditors) must	TR must be published annually on the auditor's website for the 12 months commencing on 1 July, within four months after the end of the reporting year. The first transparency report is required for the year ending 30 June 2013. The report must be	The report must include:if the auditor belongs to a network, a description of: • the network • the legal arrangements of the network • the structural arrangements of the network • a description of the auditor's internal quality control system

	publish a transparency report on their website if they have conducted audits under Div 3 of Pt 2M.3 of any combination of 10 or more of the following bodies in the past reporting year (i.e. the 12 months commencing on 1 July): listed companies listed registered schemes authorised deposit-taking institutions, as defined in section 9 of the Banking Act 1959 bodies mentioned in section 3(2)(c) or 3(2)(e) of the Australian Prudential Regulation Authority Act 1998 any other bodies prescribed by the regulations for this purpose.	lodged with ASIC before being published on the auditor's website.	• a statement that sets out the auditor's independence practices in the relevant reporting year • the name of each body that is authorised to review the auditor (e.g. ASIC or a professional accounting body) and the date of the most recent review of the auditor conducted by the body • the names of the relevant bodies in section 322 of the Corporations Act for which the auditor conducted an audit under Div 3 of Pt 2M.3 in the relevant reporting year • financial information for the auditor that relates to the relevant reporting year, including: • total revenue • revenue relating to audits of financial statements conducted by the auditor and other services provided by the auditor. For audit firms or authorised audit companies, the report must also include: • a description of the firm or company's: • legal structure • ownership • governance structure • a statement by the firm's administrative body or management body (or the company's board of directors) on the effectiveness of the functioning of the internal quality control system in the relevant reporting year • the date on which the firm or company most recently conducted an internal review of its independence compliance • a statement about the firm or company's policy on the minimum amount and nature of continuing or other professional education that professional members of an audit team must undertake during the relevant reporting year • information about the basis for remuneration of the firm's partners or the company's directors. Additional information that may be included in a transparency report is also suggested by ASIC (2013), such as AQIs, information on inspections, findings from external reviews, etc.

	If a partnership is part of a network of affiliated audit firms, the requirement applies to each individual partnership (i.e. each partnership must prepare and lodge a separate transparency report).		
US	The Centre for Audit Quality (CAQ) of the American Institute of Certified Public Accountants (AICPA) in 2014 recommended increased disclosures and proposed a voluntary model of TR for audit firms, which is an updated version of the model published in 2010 (Center for Audit Quality, 2013; AICPA, 2010 and 2014; Shong et al., 2015)	TR are voluntary. There are no TR legal dispositions. Audit firms may publish their TR in their webpages.	The latest document is a concept release on AQI (PCAOB 2015) which asks for feedback on 28 potential indicators on the fields of audit professionals, audit process and audit results

TABLE 2
Extract from document-term matrix of TR

	Number of times each term is repeated					
TR	beyond	big	black	board	build	busi
IR EY	0	0	0	0	0	0
IR Deloitte	0	0	0	0	0	0
IR KPMG	0	0	0	0	0	0
IR PwC	0	0	0	0	0	0
UK EY	0	0	0	1	0	0
UK Deloitte	1	0	0	0	0	0
UK KPMG	0	1	0	0	1	3
UK PwC	0	0	0	1	0	1
LU EY	0	0	0	0	0	0
LU Deloitte	0	0	0	0	0	0
LU KPMG	0	0	0	0	0	0
LU PwC	0	0	0	0	0	0
AT EY	0	0	0	0	0	0
AT Deloitte	0	0	0	0	0	0
AT KPMG	0	0	0	0	0	0
AT PwC	0	0	0	0	0	0
US EY	0	0	0	0	0	0
US Deloitte	0	0	0	0	0	0
US KPMG	0	0	0	0	0	0
US PwC	0	0	0	0	0	0
HU EY	0	0	0	0	0	0
HU Deloitte	0	0	0	0	0	0
HU KPMG	0	0	0	0	0	0
HU PwC	0	0	0	0	0	0
MT EY	0	0	0	0	0	0
MT Deloitte	0	0	1	0	0	0
MT KPMG	0	0	1	0	0	0
MT PwC	0	0	0	0	0	0

Note this is just an extract showing only some of the reports and some of the words in the document-term matrix. The TR are labelled with an acronym for the country plus the acronym of the firm- i.e., United Kingdom UK, Ireland IR, Luxemburg, LU, Hungary HU, Malta MT, the United Stated of America US and Australia AT, whereas EY, KPMG, Deloitte and PwC are the so called Big-4.

TABLE 3
Kruskal Wallis and Multiple Comparisons of words

Unigram sequence analysis

			Mean				K-W Signif.	Significant paired differences
			Del	EY	Kpm	PwC		
1	acceptance		4.714	10.286	7.429	4.286	***	EY-PwC; EY-Del
2	afs****	PWC Authorisation for Services system	0	0	0	0,286	.	PwC-Del; PwC-Kpm; PwC-EY
3	aqr	Audit quality review	2.286	14.714	2	2.429	**	EY-Kpm; EY-Del; EY-PwC
4	aura****	PWC global Enterprise Resource Planning system	0.143	0	0	4.714	.	PwC-Kpm; PwC-EY; PwC-Del
5	bribery		0.143	4.286	0.286	0.714	***	EY-Del; EY-Kpm; EY-PwC
6	ces	Central Entity Service with information about clients	0	0	0.143	1.000	**	PwC-Del, PwC-EY; PwC-Kpm
7	continuance		3	13.143	5.429	2.571	***	EY-PwC; EY-Del; EY-Kpm, Kpm-PwC
8	desc****	Deloitte Entity Search and Compliance System	3.429	0	0	0	.	Del-Kpm; Del-EY; Del-PwC
9	inspection		11.000	18.429	6.429	5.714	*	EY-PwC;EY-Kpm
10	rotation		2	13.714	9.143	3	***	EY-Del; EY-PwC; Kpm-Del; Kpm-PwC
11	misconduct		0	1.429	0	0.143	***	EY-Kpm; EY-Del; EY-PwC
12	ppd	Professional Practice Director	0.429	12.857	0	0	***	EY-Kpm; EY-Del; EY-PwC
13	eaudit****	KPMG's electronic audit workflow	0	0	5.429	0	***	Kpm-PwC; Kpm-Del; Kpm-EY
14	eqc****	Engagement Quality Control	0	0	7.143	0	***	Kpm-PwC; Kpm-Del; Kpm-EY
15	gcr****	Global Compliance Review	0	0	2.571	0	***	Kpm-Del; Kpm-PwC; Kpm-EY
16	gims****	Global Independence Monitoring System	4.000	0	0	0	.	Del-Kpm; Del-EY; Del-PwC

17	hotline		0.143	4.286	3.143	0	***	Kpm-PwC; EY-PwC; Kpm-Del; EY-Del
18	iesba	International Ethics Standards Board for Accountants	0.143	4.143	5.714	1.286	**	Kpm-Del; Kpm-PwC; EY-Del; EY-PwC
19	kam****	KPMG's International Audit Manual	0	0	5.143	0	***	Kpm-PwC; Kpm-Del; Kpm-EY
20	qmr****	Quality Management Review	0	0	0	2.000	*	PwC-Del; PwC-Kpm; PwC-EY
21	qpr****	Quality Performance Review	0	0	6.571	0	***	Kpm-PwC; Kpm-Del; Kpm-EY
22	rcp****	Risk Compliance Program	0	0	4.714	0	***	Kpm-PwC; Kpm-Del; Kpm-EY
23	risk		38.143	44.286	84.857	34.714	.	Kpm-Del
24	specialist		1.429	0.429	3.857	2.571	*	Kpm-EY; EY-PwC
25	industry		8.714	6.857	16.143	6.857	.	Kpm-PwC; Kpm-EY; Kpm-Del
26	judgement		1.857	0	7.143	1.857	*	Kpm-EY;
27	framework		5.429	3.429	12.143	4.143	*	Kpm-EY; Kpm-Del; Kpm-PwC
28	kics****	KPMG's web-based independence tracking system	0	0	1.857	0	*	Kpm-PwC; Kpm-Del; Kpm-EY
29	sentineltm****	KPMG's global conflicts and independence checking system	0	0	1.857	0	*	Kpm-PwC; Kpm-Del; Kpm-EY
30	global		17.286	117.714	75.143	26.714	***	Kpm-Del; EY-Del; Kpm-PwC; EY-PwC; EY-Kpm

Bigram sequence analysis

		Mean				K-W Signif.	Significant paired differences
		Del	EY	Kpm	PwC		
31	client acceptance	0.857	6.143	0.459	1.286	***	EY-Kpm; EY-Del; EY-PwC
32	continuance policy	0	1.571	0	0	***	EY-Del; EY-Kpm; EY-PwC
33	global compact	0	0.429	0	0	*	EY-Kpm; EY-Del; EY-PwC
34	tone top	0.857	2.714	9.286	0.714	**	Kpm-Del; Kpm-PwC; EY-Del; EY-PwC

Notes: Test of significance: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' Significant paired differences computed with Dunn's tests.
**** These terms are firm-specific, which explains the significant paired differences.

TABLE 4

Confusion matrix of the TRs, using extreme gradient boosting

Actual		Predicted				
		Audit firm				
		Del	EY	KPMG	PwC	Accuracy
Audit firm	Del	7	0	0	0	100.00%
	EY	0	7	0	0	100.00%
	KPMG	0	0	7	0	100.00%
	PwC	0	0	0	7	100.00%
Total						100.00%

TABLE 5
Descriptive statistics from the TR- content analysis

Variable	Definition	mean	sd	min	max	range
pages	Number of pages in TR	38.79	25.96	10.0	144.00	134.00
words	Number of words in TR	16117	13638	3733	75151.00	71418.00
photo	Reports pictures of partners (dummy var)	0.82	0.39	0.0	1.00	1.00
att_meet	Table with meeting attendance	0.14	0.36	0.0	1.00	1.00
letter_num	Number of letters presenting the TR	1.57	1.20	0.0	6.00	6.00
key_fin_data	Reports key financials (dummy var)	0.36	0.49	0.0	1.00	1.00
legal_par	Number of paragraphs on the legal structure	33.32	23.36	12.0	121.00	109.00
partners	Reports the number of partners (dummy var)	0.43	0.50	0.0	1.00	1.00
num_partners	Number of partners	244.25	306.38	5.0	885.00	880.00
QC_par	Number of paragraphs on quality control	86.29	65.82	5.0	268.00	263.00
QC_date	Reports the date of the last quality control (dummy var)	0.89	0.31	0.0	1.00	1.00
num_PIE	Number of PIE	141.64	134.61	10.0	444.00	434.00
yrs_rot	States the years of their rotation policy (dummy var)	0.39	0.50	0.0	1.00	1.00
Edu_par	Number of paragraphs on education policy	8.46	5.69	1.0	26.00	25.00
Edu_hrs	Reports the number hours of education (dummy var)	0.36	0.49	0.0	1.00	1.00
pay_par	Number of paragraphs on partners remuneration policy	5.89	3.64	2.0	20.00	18.00
audit_tot	Audit fees/ total fees	0.38	0.16	0.2	0.83	0.63
client_dis	Splits revenues per type of client (audit client vs non-audit client)(dummy var)	0.36	0.49	0.0	1.00	1.00
comparat	Reports the comparative amount of fees for the preceding year (dummy var)	0.75	0.44	0.0	1.00	1.00
fee_dep	Explicitly acknowledges that there is no fee dependence from any client (dummy var)	0.29	0.46	0.0	1.00	1.00
more_client	Explicitly acknowledges that audit partners or members of the audit team cannot be compensated for the sale of non-audit services to their audit clients (dummy var)	0.79	0.42	0.0	1.00	1.00
gender	Reports statistics of gender amongst employees (dummy var)	0.21	0.42	0.0	1.00	1.00
Ethnic	Reports statistics of ethnic origin amongst employees (dummy var)	0.18	0.39	0.0	1.00	1.00
life_work	Explains lifeworking balance policy (dummy var)	0.14	0.36	0.0	1.00	1.00

communi	Explains community engagement initiatives (dummy var)	0.36	0.49	0.0	1.00	1.00
env	Explains environmentally friendly initiatives (dummy var)	0.39	0.50	0.0	1.00	1.00
Global_comp	Has signed the Global Compact (dummy var)	0.14	0.36	0.0	1.00	1.00
firm_code	Follows the firm's Ethics code (dummy var)	0.61	0.50	0.0	1.00	1.00
iesba_code	Follows the IESBA Ethics code(dummy var)	0.79	0.42	0.0	1.00	1.00
sec	Refers to the SEC (dummy var)	0.79	0.42	0.0	1.00	1.00
pcaob	Refers to the PCAOB (dummy var)	0.79	0.42	0.0	1.00	1.00
values	States the firm's vision and values (dummy var)	0.54	0.51	0.0	1.00	1.00
tone_top	Refers to tone at the top (dummy var)	0.96	0.19	0.0	1.00	1.00
isqc	Refers to the standard of audit quality control (dummy var)	0.71	0.46	0.0	1.00	1.00
stake	Refers to communication with stakeholders (dummy var)	0.32	0.48	0.0	1.00	1.00
sanctio	Reports on sanctions received (dummy var)	0.11	0.31	0.0	1.00	1.00
awards	Reports on awards received (dummy var)	0.18	0.39	0.0	1.00	1.00
kpi	Reports key performance indicators (dummy var)	0.11	0.31	0.0	1.00	1.00
quality_par	Number of paragraphs on audit quality policy	20.36	14.92	0.0	68.00	68.00
engage_par	Number of paragraphs on acceptance engagement policy	6.79	3.89	0.0	12.00	12.00
ext_audit	Reports the external auditor (dummy var)	0.11	0.31	0.0	1.00	1.00

TABLE 6
ANOVA and Multiple Comparisons from the TR- content analysis

	FIRM ANALYSIS						COUNTRY ANALYSIS								
	Mean				ANOVA Sign.	Sign. paired differences	Mean							ANOVA Signif.	Sign. paired differences
	Del	EY	Kpm	PwC			AT	HU	IR	LU	MT	UK	US		
pages	36.29	33.43	51.43	34.00			38.75	22.25	37.50	23.00	46.50	83.50	20.00	***	UK-US, UK-HU, UK-LU, UK-IR, UK-AT, UK-MT
words	12038.86	14253.29	24545.57	13631.71			15178.00	8718.75	16369.25	9022.50	14679.50	40317.25	8536.25	**	UK-US, UK-HU, UK-LU, UK-IR, UK-AT, UK-MT
letter_num	1.14	1.86	2.14	1.14			1.75	1.00	2.00	1.00	1.75	2.50	1.00		
legal_par	27.43	33.86	34.29	37.71			19.75	28.00	42.00	24.00	23.00	74.00	22.50	**	UK-US, UK-HU, UK-LU, UK-AT, UK-MT
QC_par	52.43	87.57	138.14	67.00	.	KPMG-De	101.00	75.00	70.25	78.75	54.50	160.25	64.25		
num_PIE	123.57	123.14	170.00	149.86			319.50	19.75	213.75	89.50	37.00	296.25	15.75	***	IR-US, UK-US, AT-US, IR-HU, UK-HU, AT-HU, IR-MT, UK-MT, AT-MT, IR-LU, UK-LU, AT-LU
Edu_par	6.00	8.00	9.43	10.43			13.50	4.00	7.50	8.25	7.25	13.00	5.75		
pgay_par	7.29	7.00	5.00	4.29			4.75	4.00	6.00	4.50	5.00	11.75	5.25	*	UK-HU, UK-LU, UK-AT, UK-MT, UK-US
audit_tot	0.32	0.37	0.41	0.42			0.26	0.68	0.39	0.38	0.35	0.23	0.34	***	HU-UK, HU-AT, HU-US, HU-MT, HU-LU, HU-IR
quality_par	15.14	13.00	31.00	22.29	.	KPMG-EY	24.75	14.50	23.00	21.00	13.25	34.75	11.25		
engage_par	5.57	9.29	9.29	3.00	**	EY-PWC KPMG-PWC	6.00	6.75	6.00	7.00	7.25	8.50	6.00		

Notes: Test of significance: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1. Significant paired differences computed with Tukey HSD tests. Variables are defined in Table 5.

TABLE 7
Contingency test from the TR- content analysis

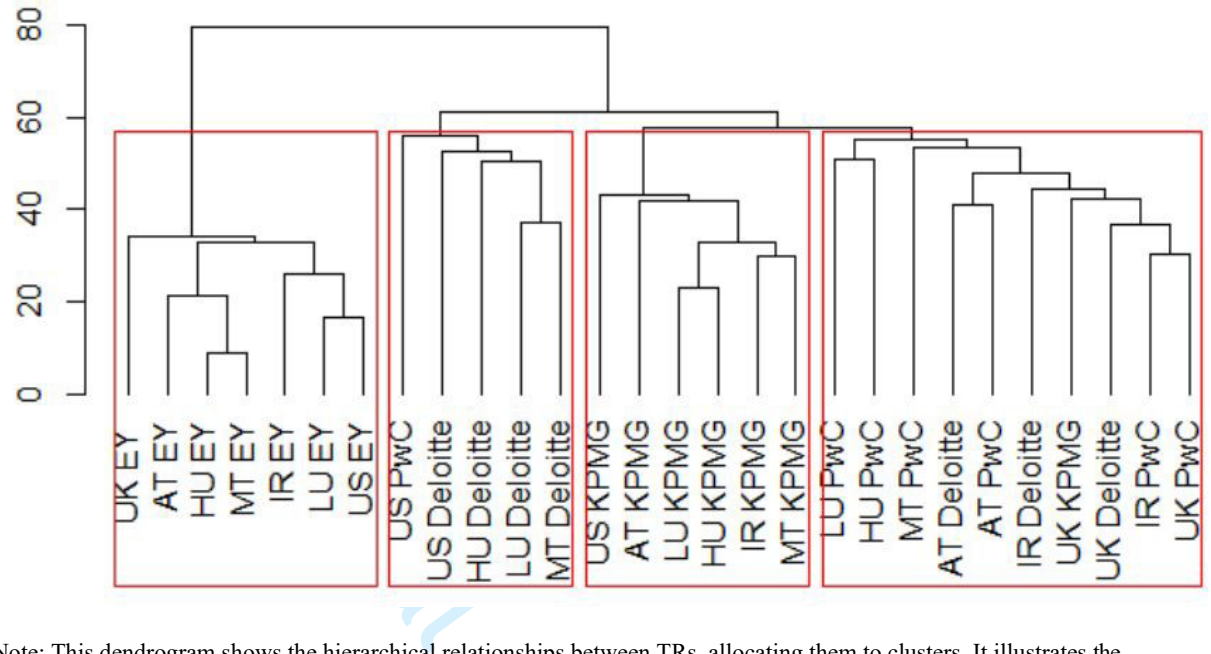
	FIRM ANALYSIS					COUNTRY ANALYSIS							
	Mean				Fisher Test Signif.	Mean							Fisher Test Signif.
	Del	EY	Kpm	PwC		AT	HU	IR	LU	MT	UK	US	
Photo	0.57	1.00	1.00	0.71	0.120	1.00	0.50	1.00	0.75	1.00	1.00	0.50	0.234
att_meet	0.29	0.00	0.00	0.29	0.279	0.00	0.00	0.50	0.00	0.00	0.50	0.00	0.070
key_fin_data	0.29	0.29	0.43	0.43	1.000	0.75	0.00	0.00	0.00	0.50	1.00	0.25	0.006 **
partners	0.14	0.57	0.57	0.43	0.361	0.50	0.00	0.75	0.25	0.75	0.75	0.00	0.102
QC_date	0.71	1.00	0.86	1.00	0.581	0.75	1.00	1.00	0.50	1.00	1.00	1.00	0.316
yrs_rot	0.43	0.71	0.00	0.43	0.060 .	0.50	0.25	0.50	0.25	0.25	0.25	0.75	0.864
Edu_hrs	0.29	1.00	0.14	0.00	0.000 ***	0.50	0.25	0.25	0.25	0.50	0.25	0.50	1.000
client_dis	0.43	0.14	0.29	0.57	0.562	0.75	0.00	0.25	0.00	0.50	1.00	0.00	0.006 **
comparat	0.86	0.71	0.71	0.71	1.000	1.00	0.50	1.00	0.25	1.00	1.00	0.50	0.030 *
fee_dep	0.14	0.00	0.86	0.14	0.001 ***	0.25	0.25	0.25	0.25	0.50	0.25	0.25	1.000
more_client	0.57	1.00	0.71	0.86	0.430	1.00	0.25	1.00	1.00	1.00	0.50	0.75	0.059 .
Gender	0.14	0.00	0.29	0.43	0.430	0.00	0.00	0.25	0.00	0.75	0.50	0.00	0.059 .
Ethnic	0.14	0.00	0.29	0.29	0.706	0.00	0.00	0.25	0.00	0.50	0.50	0.00	0.234
life_work	0.29	0.00	0.14	0.14	0.882	0.00	0.00	0.25	0.00	0.75	0.00	0.00	0.033 *
communi	0.29	0.43	0.43	0.29	1.000	0.25	0.25	0.75	0.00	0.75	0.50	0.00	0.154
Env	0.14	0.57	0.43	0.43	0.530	0.50	0.25	0.75	0.00	0.75	0.50	0.00	0.179
Global_comp	0.00	0.57	0.00	0.00	0.006 **	0.25	0.25	0.00	0.00	0.25	0.25	0.00	1.000
iesba_code	0.29	1.00	0.86	1.00	0.004 **	0.75	0.75	0.75	1.00	1.00	0.75	0.50	0.923
Sec	0.57	1.00	1.00	0.57	0.047 *	0.75	0.75	1.00	0.50	0.50	1.00	1.00	0.495
Pcaob	0.71	0.86	1.00	0.57	0.430	0.75	0.50	1.00	1.00	0.25	1.00	1.00	0.059 .
Values	0.29	1.00	0.71	0.14	0.003 **	0.50	0.50	0.75	0.50	0.75	0.50	0.25	0.965
tone_top	0.86	1.00	1.00	1.00	1.000	1.00	0.75	1.00	1.00	1.00	1.00	1.00	1.000
Isqc	0.00	1.00	0.86	1.00	0.000 ***	0.75	0.75	0.75	0.75	0.75	0.75	0.50	1.000
Stake	0.43	0.57	0.29	0.00	0.182	0.25	0.25	0.25	0.00	0.75	0.75	0.00	0.115
Sanctio	0.00	0.14	0.14	0.14	1.000	0.00	0.00	0.00	0.00	0.00	0.75	0.00	0.008 **
awards	0.29	0.14	0.14	0.14	1.000	0.00	0.00	0.50	0.00	0.25	0.50	0.00	0.234
KPI	0.00	0.00	0.14	0.29	0.581	0.00	0.00	0.25	0.25	0.00	0.25	0.00	1.000
ext_audit	0.00	0.14	0.14	0.14	1.000	0.00	0.00	0.00	0.00	0.00	0.75	0.00	0.008 **

Notes: Test of significance: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1. Variables are defined in Table 5.

FIGURE 1
Word cloud of terms occurring most frequently in all TRs

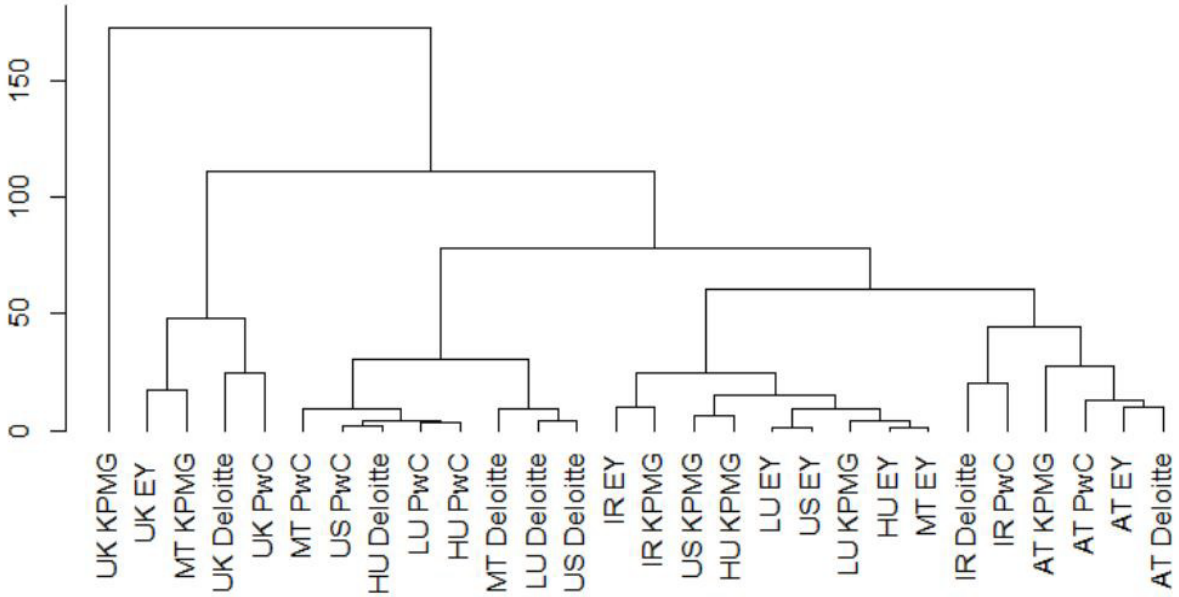
Note: This figure depicts word stems, not words.

FIGURE 2
Dendrogram of TRs from weighted TF-IDF bag-of-words



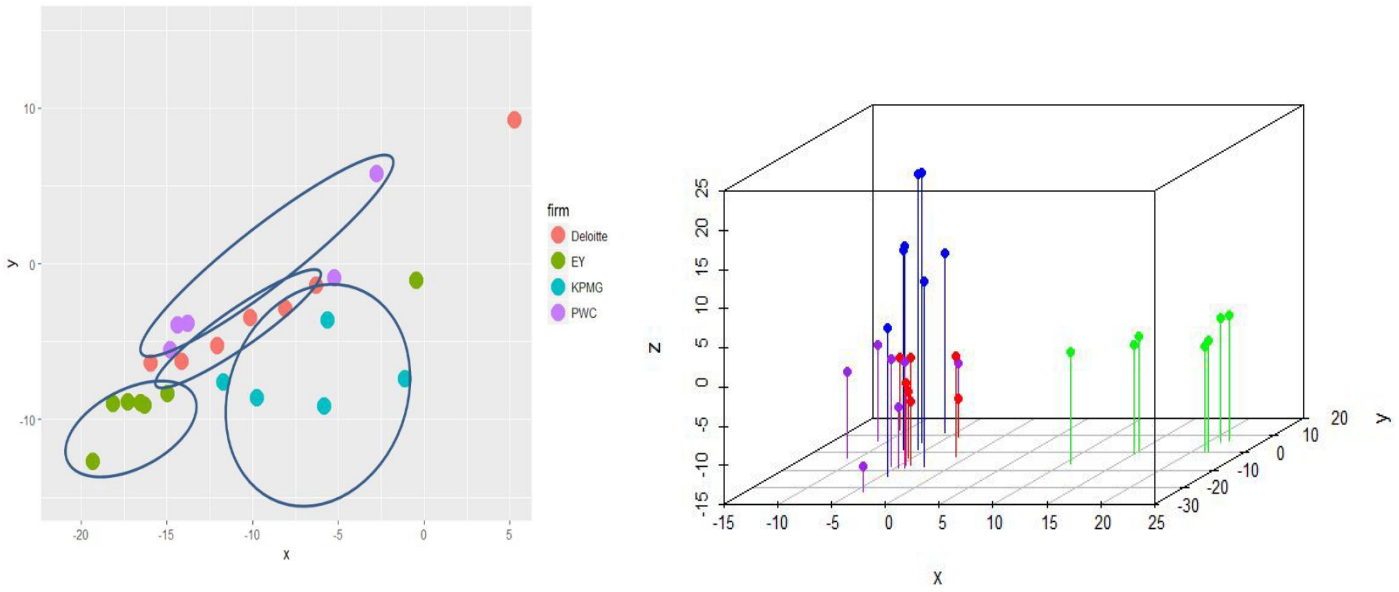
Note: This dendrogram shows the hierarchical relationships between TRs, allocating them to clusters. It illustrates the hierarchical organization of TRs from weighted TF-IDF bag-of-words.

FIGURE 3
Dendrogram of TRs from the latent semantic space



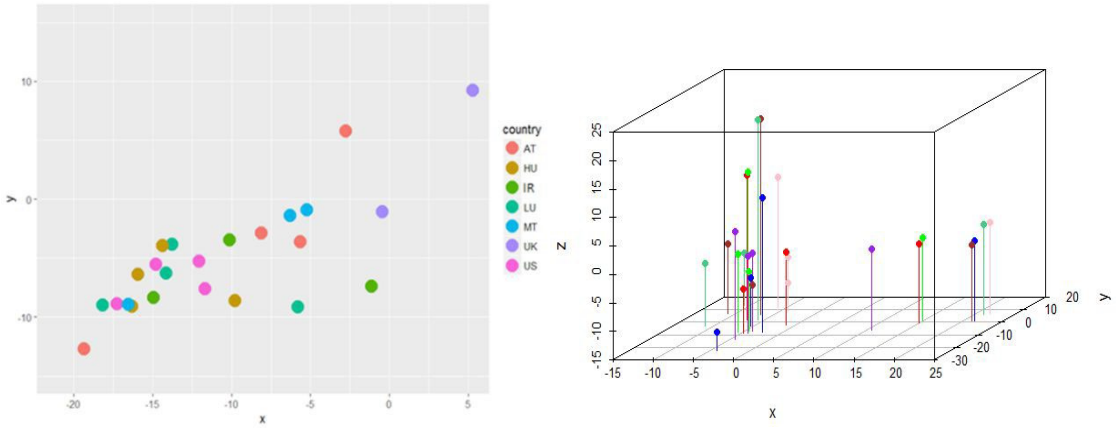
Note: This dendrogram shows the hierarchical relationships between TRs, allocating them to clusters. It illustrates the hierarchical organization of TRs from the latent semantic space.

FIGURE 4
2-D and 3-D representations from the latent semantic space, colored by audit firm



Note: These figures depict the latent semantic space in two and three dimensions respectively. Each point represents a TR, colored depending on the Big-4 signing the report.

FIGURE 5
2-D and 3-D representations from the latent semantic space, colored by country



Note: These figures depict the latent semantic space in two and three dimensions respectively. Each point represents a TR, colored depending on the country of the report.

FIGURE 6

Correlogram plot of TRs from the weighted TF-IDF bag-of-words

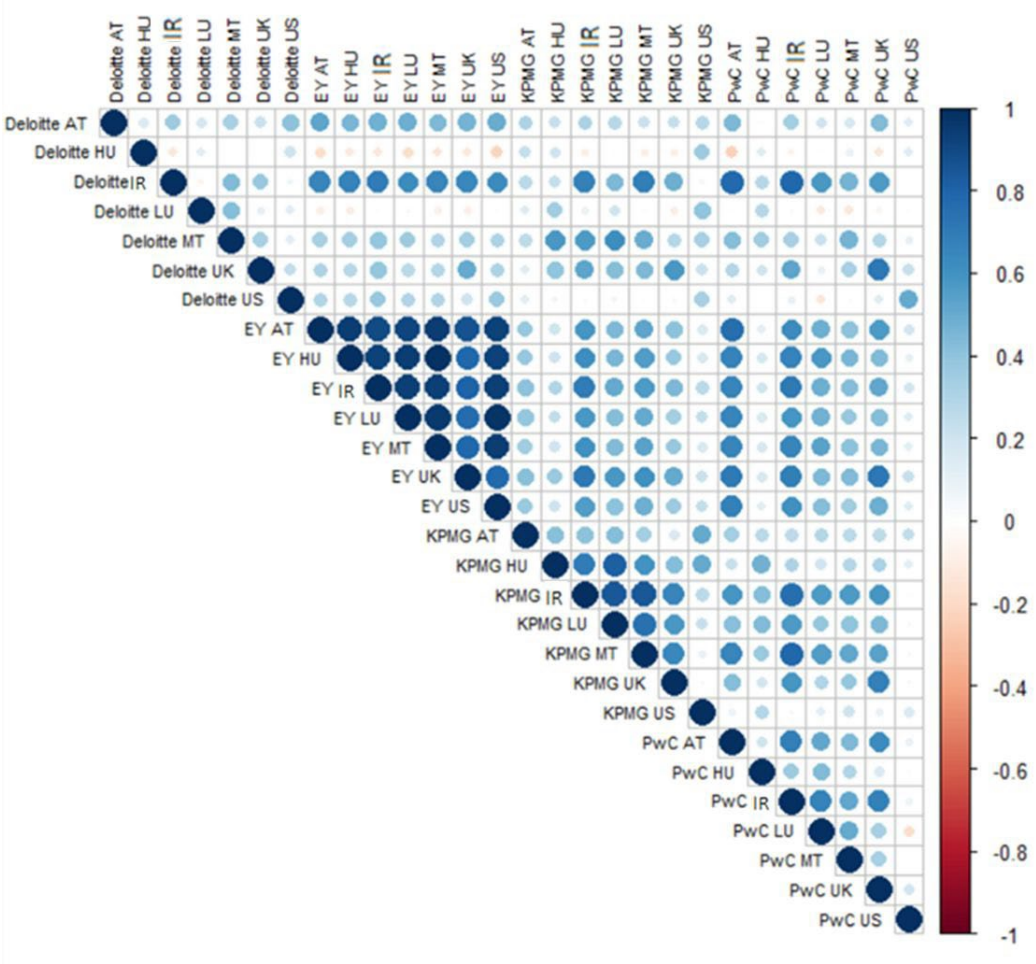
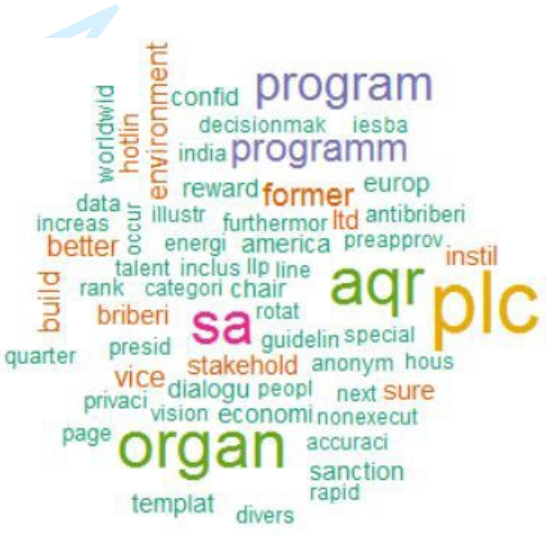


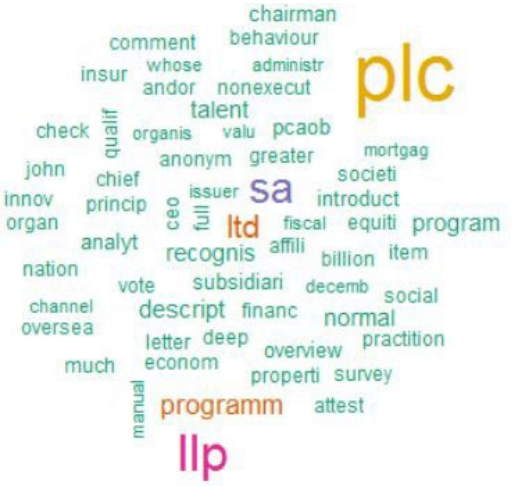
FIGURE 7

Word cloud of terms occurring most frequently the TR by Big-4

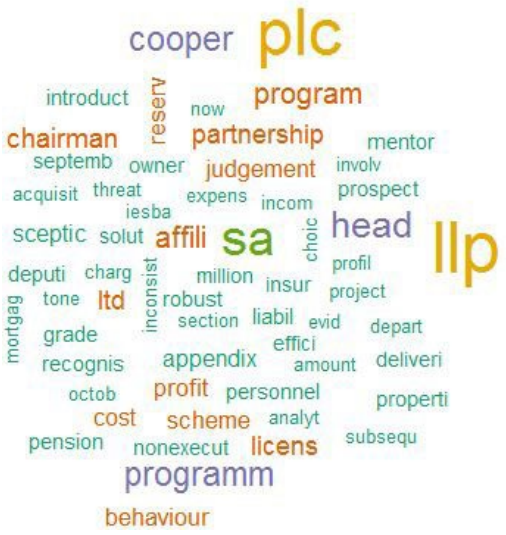
EY



Deloitte

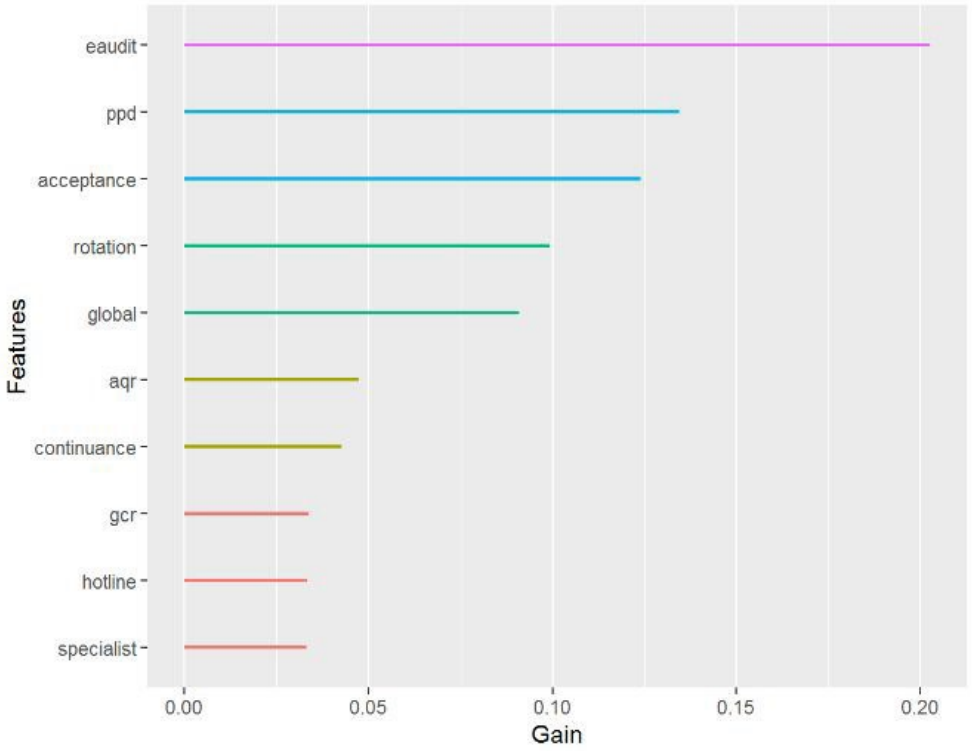


KPMG



PWC

FIGURE 8
Feature importance in Big 4 TR classification using extreme gradient boosting



Notes: Variables are defined in Table 3.