International Journal of Youth Entrepreneurship Studies
WISDOM ACADEMIC https://doi.org/10.63808/488.v2i2.488 Volume 2 , Issue 2

Review Article

Do Key Audit Matters Really Inform? Boilerplate Language, Textual Similarity, and Investor Reaction

Guangwei Hao[1,*]
[1]Central University of Finance and Economics, Beijing 100081, China
[*]**Correspondence:** Guangwei Hao, 174238223@qq.com

Abstract
Auditing standards stipulate that KAMs should be specifically relevant to the situation at hand and not standardized in their wording, archival research examining the extent to which KAMs are communicated to investors has been inconsistent within and across different countries, sample years, and study approaches. The current review classifies fifteen pieces into fourteen articles based on how KAM disclosures were quantified rather than on results obtained, distinguishing presence versus magnitude, generic versus specific, and textual similarity. Work done at the coarsest level is consistent on non-existence but divided about what should be measured, whereas textual measures lead to similar conclusions: it is specificity that determines the link between disclosure and market reaction, while similarity moderates the extent of this relationship but does not have any impact on its own. Consistency of results follows the level of disclosure measurement. The number of boilerplates used in the UK increased almost twice during the post-reform era, whereas European issuers gathered boilerplates and specificity together.

Keywords
Key audit matters, Boilerplate language, Textual similarity, Investor reaction, Auditor reporting

1. Introduction

The auditing standard requires something unique. The key audit matter should be articulated in such a way that the account becomes related to the situation of the audited organization, and no general wording should be used; China's CSA 1504 published in 2016 and implemented over 2017-2018, states this in almost identical terms to other regulators (CICPA 2016).

Whether there has been informative disclosure from the requirement is a separate issue altogether, and there is no convergence on this matter as well. Reviewing 49 empirical studies shows that reactions of the stakeholders have been inconsistent, especially in the shareholder studies (Velte and Issa 2019). Archive studies regarding the UK regime, with observations ranging from 1,208 to 1,320 firms-years from 2011-2015, found that there was no difference in abnormal returns and trading volumes post-implementation of the extended report (Gutierrez et al. 2018). Research conducted on the same regulatory regime yielded different findings in the case of the measure being textual, with KAM statements containing more client-specific information being positively correlated with cumulative abnormal returns and abnormal trading volume on the date of the filing (Seebeck and Kaya 2023). Same jurisdiction, same reform, different results.

The distinction is typically understood as a disagreement on the issue of the informative nature of KAMs. The disagreement stems from another source entirely. The dividing line between the two groups of results does not lie in the nation under study, the time period under consideration, or the methodology employed, but rather the level of analysis of the disclosures in question: presence and frequency in the former, text characteristics in the latter.

The regrouping of literature addresses a further, minor issue. Boilerplate and specificity are regularly considered as opposing ends of a continuum. This is not the case: boilerplate is measured relative to a corpus of other studies, while specificity is measured relative to the study itself, and they have grown together in European patent documents. Once again, similarity is a separate measurement, one that is calculated based on two documents. The number of studies is not novel; it is only the way they are grouped that is so.

2. The regulatory promise and the divided evidence

The three reporting frameworks have one thing in common: one single requirement. The Financial Reporting Council of the UK amended its auditor reporting framework in 2013 such that auditors of premium listed entities were required to provide information on the risk of material misstatement that had the biggest impact on the audit. ISA 701 came up in 2015, while the PCAOB issued critical audit matters in 2017. However, there are differences in the thresholds and scope of each of them; in fact, the US filter includes an extra criterion for consideration. The requirement in itself does not change much in its wording. They all operate as if generic language meant lack of information. The two factors being bundled together are actually not necessarily connected to one another: how far the passage is from the rest of the reporting and how much it says about the audited firm.

Fourteen studies in the archival literature provide the foundation for the following analysis, having been published from 2018 to 2026 and representing data from the UK, US, Germany, China, Taiwan, and the aggregate of Europe. Outcome variables do not remain uniform across these studies. Some studies analyze price effects while others analyze volume and neither in the remaining ones. Table 1 provides this information on a study-to-study basis.

Research at this stage is differentiated based on the variables of interest in their outcomes rather than on the results themselves. In cases where abnormal return and trading activity were the dependent variables, consensus was reached, as there was no response generated by the adoption in the UK (Gutierrez et al. 2018), which was later confirmed by a study that revealed that the lack of reaction was caused by prior learning by investors of the risks involved (Lennox et al. 2023); and in critical audit matter disclosures in the USA (Burke et al. 2023). The papers that identify effects measure something else altogether — greater idiosyncratic risk, beta, and cost of equity with changes in spreads and forecasting uncertainty in one paper (Elsayed et al. 2023), enhanced reporting quality without any commensurate increase in audit fees in the other (Reid et al. 2019). Being associated with risk fundamentals and improving reporting quality are not aspects of the news, and a news report could provide one of these without the other. The group cannot even agree upon what the question is at hand, such that both instrument and target are intertwined.

There is one comparison, however, which gives rise to divergent results regarding the same outcomes. Gutierrez et al. (2018) and Seebeck and Kaya (2023) investigate filers from the UK across the same period and analyze abnormal return and abnormal trading volume surrounding the filing date. The only difference between them is that while the former notes whether an extended disclosure was made and how much of it, the latter observes what the disclosure contained. If this is the source of divergence between them, then finer studies should have reacted differently.

3. Two textual measures and investor reaction

3.1. Boilerplate and specificity

The concept of boilerplate is quantified by finding four words that appear repeatedly in a majority of the reports belonging to the corpus and counting their occurrence in each section of interest according to the definition of boilerplate, which is wording that appears repeatedly in filings for a particular year (Lang and Stice-Lawrence 2015). This measure is word order-sensitive, in that four words in different order make up a different sequence. The specificity measure relies on Named Entity Recognition based on Stanford, where persons, locations, organizations, percentages, financial values, and dates are detected, and their presence in the sequence is indicated (Seebeck and Kaya 2023).

The evolution of both these characteristics after the reforms has been recorded in a number of markets, but the patterns are different. The number of boilerplate four-grams in UK KAMs increased from around 34 in the first year to around 66 in the sixth year post-reform, whereas the average number of matters per report decreased from 3.98 to around 3.83 (Seebeck 2024). For European companies, KAM disclosures from 2016 onward have become more lengthy, more quantitatively oriented, and more specific, but also more boilerplate-oriented (Küster 2024). In contrast, the number and length of KAM disclosures have decreased in the United States in the four years since the new rules took effect (Calabrese et al. 2026).

The European model deserves consideration, as greater specificity and boilerplate emerging at the same time is hard to square with considering them opposites on a spectrum. The processes involved make this clear. Boilerplate is calculated relative to a database of other reports; specificity is calculated relative to the report itself, as it counts the names it uses. No restrictions exist for a passage to score well in both aspects. The concept of similarity, which refers to the computation of similarity between two documents, is another different concept; this explains why textual similarity between two reports does not automatically mean that one of the two reports has moved towards common language use (Chang et al. 2026). KAM, which quotes an exact figure for the savings of households in the country, is specific but useful to any client in the industry.

If the market reacts, it is reacting to the latter rather than the former. According to studies from the UK, high specificity in KAM reports was associated with higher cumulative abnormal returns and high abnormal trading volumes during the announcement period while proxies for communicative value from the same set of KAM reports, namely readability, evaluative content, and visual elements, had no association (Seebeck and Kaya 2023). Filings from Taiwanese respondents arrive at the same result via another path. KAMs containing more specific client information relate to lower reporting quality perceptions, in particular due to the risk description rather than the risk response, and the same textual variable predicts the occurrence of abnormal accruals and misstatements (Chang, Chi, and Stone 2024). The two findings diverge in sign but concur on the issue at hand, for whereas one documents the size of a response, the other captures the direction thereof, and both capture the impact of specificity on the market in a way that presence and counts do not. The fact that the indicator is predictive of actual results further limits one question a correlation alone cannot answer, given that a proxy in place of uncorrelated variance has no cause to predict errors. This is where the primary issue resides.

3.2. Textual similarity

The measures below contribute conditionally rather than directly.

Similarity is assessed along two dimensions. One dimension concerns persistence, or to what extent this year's report uses similar language as last year's for the same client (Brown and Tucker 2011). The other dimension concerns similarity among contemporary reports, which can be assessed through one of three benchmarks – similarity of a report relative to all other reports within a given reporting period, similarity of a report relative to other reports produced by the same audit firm, or similarity relative to other reports about the same type of risk. Firm-level resemblance points toward internal templates; risk-level resemblance may reflect only shared exposure.

Cosine similarity handles the calculations, and it is independent of the length of the documents and word order. It is precisely the second factor that distinguishes cosine similarity from the four-word sequence method since two expressions of the same meaning but using different word orders would be considered dissimilar in n-grams (Deneuve 2026).

Cross-country evidence on the evolution of similarity is based on two countries. The similarity between UK reports has converged on all measures in each year studied (Seebeck 2024). German auditors have employed similar language to disclose identical information for the same client in consecutive years if they had not switched firms (Carlé et al. 2023). The leaving company takes the template with them, and the arriving company does not have a previous report to build upon; the reporting design does not separate either. The comparative longitudinal data for US filers is not available; the study on adoption of CAM, which is cited here, considers numbers and document length but not similarity (Calabrese et al. 2026). The evidence regarding the process of convergence is entirely European.

The empirical findings for the direct market tests are inconsequential. There is no clear relationship between the above similarities and abnormal returns or abnormal trading volume, and the models do not explain much about market reaction (Seebeck 2024). It's a typical null result when read alone. When read against the results of the tests for specificity, however, it is the more fascinating result of the two because in the same regulation regime and within the same reporting period, the same set of documents generates a significant relationship in one aspect but not in another.

3.3. The interaction and the review's claim

These two characteristics do not work separately from each other. With 1,045 firm-year observations obtained from UK filers in the fiscal years from 2012 through 2018, another study found that the relationship between disclosure specificity and stock market reaction is negatively moderated by disclosure similarity for both abnormal trading volume and absolute cumulative abnormal returns (Seebeck 2024). What the result supports is a proposition available for testing rather than a settled regularity: precise information appears to be discounted when it arrives in a form investors have seen before.

Two qualifications should be added. First, the agreement between textual research is based on three criteria of investor behavior, which are derived from two countries and two different sets of authors; the rest of the textual research tells us about the movement of these properties. The interaction depends upon only one study in the UK. When placed in relation to the previous literature, this produces a structure of three components, but not in equal parts. Measures that depend upon the presence, the number, and the length of boilerplate clauses have never created a market response, and the studies belonging to this category that reported effects did so only because of the manipulation of the dependent variable, which is why it qualifies as a control case instead of a competing one. The effect of boilerplate clauses and specificity comes from the property of naming, not that of avoiding repetition.

This is not an assertion regarding whether key audit matters are informative. This is rather an assertion on the reason why there have been differing answers to this question. Consistency in findings is based on the level of disclosure analysis. Those at the highest level of analysis all agree in one absence, while those at the lower levels all agree on the property. Fifteen items in Table 1 are arranged in accordance with their type of measurement and show the outcome variables along with the direction of the effect. This arrangement allows one to see the relationship without any explanation. There is one thing in this table that can easily be overlooked. Outcome variables are the least consistent in the group that has elicited no response, which implies that both the instrument and the stimulus have moved together and not separately.

Table 1

Evidence by measurement type

Study	Country	Measurement type	Dependent variable	Direction	Role	Pub.
Gutierrez et al.	UK	Presence & scale	Returns, volume	No reaction	IV	2018
Reid et al.	UK	Presence & scale	Reporting quality, costs	Quality improved	IV	2019
Lennox et al.	UK	Presence & scale	Returns	No reaction	IV	2023
Elsayed et al.	UK	Presence & scale	Risk, cost of equity, spread	Higher risk measures	IV	2023
Burke et al.	US	Presence & scale	Returns, volume	No incremental information	IV	2023

WISDOM ACADEMIC https://doi.org/10.63808/488.v2i2.488 Volume 2 , Issue 2

Seebeck & Kaya	UK	Boilerplate & specificity	|CAR|, volume	Positive	IV	2023
Chang, Chi & Stone	Taiwan	Boilerplate & specificity	Perceived quality (ERC)	Specificity priced, negative valence	IV	2024
Küster	Europe	Boilerplate & specificity	Disclosure characteristics	Both rising	DV	2024
Calabrese et al.	US	Boilerplate & specificity	Disclosure characteristics	Count, length decline	DV	2026
Seebeck	UK	Textual similarity	|CAR|, volume	No direct effect	IV	2024
Carlé et al.	Germany	Textual similarity	Determinants	Lower after firm change	DV	2023
Deneuve	Europe	Textual similarity	Audit quality, report lag	Lower quality	IV	2026
Zeng et al.	China	Textual similarity	Audit quality	Differs by component	DV	2021
Chang et al.	Taiwan	Textual similarity	Earnings comparability	Responses only	IV	2026
Seebeck	UK	Similarity × specificity	|CAR|, volume	Attenuates specificity	IV	2024

Note: IV entries treat the measure as an explanatory variable; DV entries take textual properties as the object of explanation, documenting how these properties have moved rather than whether investors respond.

4. What is disputed and what does not transfer

These regimes depend on an assumption that has not been studied nearly as much as the disclosure regulations themselves. Standard language is seen as a sign of the lack of information, while differentiation is viewed as a sign of the presence of the information. There are two tests that verify the assumption in question. The difference in the language of European KAM disclosures is associated with low audit quality and long audit lag time – that is not what the assumption suggests (Deneuve 2026). Further proof from Taiwan's filers distinguishes between the two parts of a KAM. When auditors adopt similar approaches to dealing with an issue, the earnings reported by their clients tend to become more similar; however, the similarity of risk descriptions does not seem to be associated with this effect (Chang et al. 2026). This evidence converges in terms of one part but not the other, and viewing it as a unitary problem misses that point.

What the results reveal is a presumption that precedes the evidence that can be derived from the text. An issue that exists in every reporting period and whose description remains constant through the years is correctly represented, and the removal of any repetition of language would only result in a loss of information. Industry-wide language use could be industry-wide exposure. It has never been conclusively distinguished whether this convergence between the UK and Europe is due to template provision by the auditor or risk sharing on the client's side, as there have been no designs in literature where this point has been clarified. The correlations available fit both theories. (Carlé et al., 2023). This means that convergence is a depiction of what the writing does rather than an assessment of what the writing lacks. The hypothesis misidentifies its subject. A requirement for the writing to be different can be fulfilled by making it different, and making it different is a cost to the auditor without any benefit to the reader.

Two factors serve to restrict how much all this can generalize. There is a geographic dimension to the research data: much of what is known is based on the UK, the US, and Taiwan, with the UK doing all the work in the similarity area. This bias goes beyond sample size. If four audit firms account for virtually every listed corporation, then using a similarity score with an audit firm benchmark will leave nothing in the way of variability. And once that is the case, the firm benchmark and market benchmark begin to coincide. Trajectory results can only be taken conditionally, pending evidence from other regimes.

One possible explanation is worthy of investigation, rather than being dismissed out of hand. Textual measure studies came out after presence and count studies; hence, there is partial collinearity between measurement level and research age, as well as the availability of data that the age entails. While the comparison underlying this survey is partially shielded from confounding variables because Gutierrez et al. (2018) and Seebeck & Kaya (2023) use the same periods and files but different methods of analysis, it is not fully protected, and neither has any study been done in isolation in the literature. Here, an hypothesis is offered for testing a discrepancy in literature, not a pattern in the data itself.

5. Conclusion and future research

For anyone assessing the evidence on audit committee disclosures, there is a preliminary issue to be sorted out: At what level was the measure taken? There has been no effect of presence, frequency, and length in creating a response from the market, but the effect reported at that level came about because the dependent variable was changed. The textual measures that have generated results that hang together are based on the property that records what the report mentions, not what it does not repeat. It imposes as well a moderate expectation from any future research in this field because an article analyzing the informativeness of such disclosures must explain to its readers which textual characteristic it has selected and why it has selected this one instead of the other.

There are three such opportunities arising from the collected evidence. Chinese filings provide the most convincing of these. Language used in these is very much company specific and differs between reporting components of the same issue (Zeng et al. 2021), which is exactly the component-based framework under which conflicting results emerged in the Taiwanese context, while CSA 1504 applied to varying groups of issuers each year. An implementation strategy that has a staggered implementation on a market where there is existing variance at the component level provides an identification framework that cannot be matched by the UK approach because of its uniform implementation strategy. The second window is the horizon. Virtually all knowledge about market reaction is based on horizons of just a few days, with no knowledge of long horizons. The third is the interaction itself.

There is one limitation underlying everything said above. The fact is that each and every measure mentioned takes into account shared vocabulary, meaning that identical passages that have been expressed using different words will be treated differently, and the spelling of words can make the difference when it comes to the content. It is possible that the convergence estimate that is based on such measures underestimates or overestimates the substantive similarity of the reports. It is not known if measuring content will preserve the pattern.

Acknowledgments

The authors declare that no specific acknowledgments are required.

Author contributions

Not applicable.

Funding

This research received no external funding.

Data availability

No datasets were generated or analysed during the current study.

Declarations

Conflict of interest

The authors declare that there is no conflict of interest.

Ethics approval

Not applicable.

Consent to participate

Not applicable.

Consent for publication

All authors have given their consent.

Publisher's Note

Wisdom Academic Press Ltd. remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

[1] Brown, S. V., & Tucker, J. W. (2011). Large-sample evidence on firms' year-over-year MD&A modifications. *Journal of Accounting Research*, *49*(2), 309–346. https://doi.org/10.1111/j.1475-679X.2010.00396.x

[2] Burke, J. J., Hoitash, R., Hoitash, U., & Xiao, S. (2023). The disclosure and consequences of U.S. critical audit matters. *The Accounting Review*, *98*(2), 59–95. https://doi.org/10.2308/TAR-2021-0013

[3] Calabrese, K., Durkin, M., & Suslava, K. (2026). Changes in the reporting of critical audit matters. *International Journal of Auditing*. Advance online publication. https://doi.org/10.1111/ijau.70033

[4] Carlé, T., Pappert, N., & Quick, R. (2023). Text similarity, boilerplates and their determinants in key audit matters disclosure. *Corporate Ownership and Control*, *20*(2). https://doi.org/10.22495/cocv20i2art4

[5] Chang, Y. T., Chen, J. Z., Chen, M. H., & Chi, W. (2026). Key audit matter disclosures and earnings comparability. *Accounting and Business Research*, 1–28. Advance online publication. URL: https://nccur.lib.nccu.edu.tw/handle/140.119/162521

[6] Chang, Y. T., Chi, W., & Stone, D. N. (2024). Is client-specific information useful to investors? Evidence from key audit matter reports. *Journal of Accounting, Auditing and Finance*, *39*(3), 786–806. https://doi.org/10.1177/0148558X221117194

[7] Chinese Institute of Certified Public Accountants. (2016). *Chinese Standard on Auditing 1504: Communicating key audit matters in the independent auditor's report*. Beijing: CICPA. URL: https://www.mof.gov.cn

[8] Deneuve, E. (2026). Dissimilarity in key audit matters: Determinants and consequences. *Accounting in Europe*, *23*(1), 116–145. https://doi.org/10.1080/17449480.2025.2451382

[9] Elsayed, M., Elshandidy, T., & Ahmed, Y. (2023). Is expanded auditor reporting meaningful? UK evidence. *Journal of International Accounting, Auditing and Taxation*, *53*, 100582. https://doi.org/10.1016/j.intaccaudtax.2023.100582

[10] Gutierrez, E., Minutti-Meza, M., Tatum, K. W., & Vulcheva, M. (2018). Consequences of adopting an expanded auditor's report in the United Kingdom. *Review of Accounting Studies*, *23*(4), 1543–1587. https://doi.org/10.1007/s11142-017-9424-4

[11] Küster, S. (2024). The determinants of linguistic features in key audit matters: Empirical evidence

from Europe. *International Journal of Auditing*, *28*(3), 582–614. https://doi.org/10.1111/ijau.12344

[12] Lang, M., & Stice-Lawrence, L. (2015). Textual analysis and international financial reporting: Large sample evidence. *Journal of Accounting and Economics*, *60*(2–3), 110–135. https://doi.org/10.1016/j.jacceco.2015.09.002

[13] Lennox, C. S., Schmidt, J. J., & Thompson, A. M. (2023). Why are expanded audit reports not informative to investors? Evidence from the United Kingdom. *Review of Accounting Studies*, *28*(2), 497–532. https://doi.org/10.1007/s11142-021-09650-4

[14] Reid, L. C., Carcello, J. V., Li, C., Neal, T. L., & Francis, J. R. (2019). Impact of auditor report changes on financial reporting quality and audit costs: Evidence from the United Kingdom. *Contemporary Accounting Research*, *36*(3), 1501–1539. https://doi.org/10.1111/1911-3846.12486

[15] Seebeck, A. (2024). Back to where it started? — Do expanded auditor's reports become sticky, generic and boilerplate over time? *International Journal of Auditing*, *28*(3), 536–561. https://doi.org/10.1111/ijau.12343

[16] Seebeck, A., & Kaya, D. (2023). The power of words: An empirical analysis of the communicative value of extended auditor reports. *European Accounting Review*, *32*(5), 1185–1215. https://doi.org/10.1080/09638180.2021.2021097

[17] Velte, P., & Issa, J. (2019). The impact of key audit matter (KAM) disclosure in audit reports on stakeholders' reactions: A literature review. *Problems and Perspectives in Management*, *17*(3), 323–341. https://doi.org/10.21511/ppm.17(3).2019.26

[18] Zeng, Y., Zhang, J. H., Zhang, J., & Zhang, M. (2021). Key audit matters reports in China: Their descriptions and implications of audit quality. *Accounting Horizons*, *35*(2), 167–192. https://doi.org/10.2308/HORIZONS-19-169