

Article

Impression Management in Corporate Reporting: The Impact of MD&A Textual Similarity on CSR Outcomes

Runyu Kang¹, Ruopiao Zhang^{1,*}, Yilu Zhu²¹Department of Accounting and Finance, School of Business, Macau University of Science and Technology, Macau, China²School of Accounting, Guangzhou College of Commerce, 511363 Guangzhou, Guangdong, China*Correspondence: rpzhang@must.edu.mo (Ruopiao Zhang)

Academic Editor: Simon Jebben

Submitted: 5 August 2025 Revised: 8 December 2025 Accepted: 29 January 2026 Published: 31 July 2026

Abstract

This research explores the connection between the textual similarity of Management's Discussion and Analysis (MD&A) sections in corporate reports and subsequent Corporate Social Responsibility (CSR) performance. Using a large sample of Chinese A-share listed firms from 2010 to 2021, the study finds a significant negative relationship. Firms exhibiting higher MD&A textual similarity indicating more repetitive narrative disclosures year after year—tend to show poorer future CSR outcomes. This suggests that companies may use repetitive narratives as a passive form of impression management to conceal weak CSR engagement. The analysis further shows that this effect is mediated by increased information asymmetry, proxied by analyst forecast dispersion. Repetitive disclosures create a less transparent information environment, weakening external monitoring and giving managers greater scope to reduce investment in substantive CSR activities. The study's conclusions are supported by multiple robust empirical strategies, including instrumental variables and dynamic panel Generalized Method of Moments (GMM) estimations. These findings underscore the importance for stakeholders of scrutinizing narrative consistency in corporate reports as a potential indicator of superficial CSR efforts and a lack of genuine corporate accountability.

Keywords: Management's Discussion and Analysis (MD&A); corporate social responsibility (CSR); corporate sustainability; impression management

JEL: M41, G34, Q56, M14

1. Introduction

Strategic corporate communication plays a pivotal role in shaping perceptions of legitimacy and trustworthiness (Du and Yu, 2021). Financial reports, especially voluntary social and environmental disclosures, serve as a significant platform for managerial narratives that can greatly influence stakeholders (Deegan, 2002). However, impression management, a phenomenon long recognized in social psychology, infiltrates corporate reporting, necessitating critical analysis of how organizations portray their financial, environmental, and governance complexities (Merkel-Davies and Brennan, 2007). Through impression management strategies, executives may selectively shape disclosures to highlight favorable aspects of corporate performance, while obscuring deficiencies.

The integration of methodologies from social psychology, known as 'vertical borrowing' (Highhouse et al., 2009), has revitalized the study of impression management in corporate reporting. A key focus of this work is the Management's Discussion and Analysis (MD&A) section of annual reports, which goes beyond numerical datasets to provide narrative context on an organization's past performance and future prospects (Sonpar and Golden-Biddle, 2008; Sun, 2010; Tam et al., 2023; Muslu et al., 2015; Zhang et al., 2023). Prior studies demonstrate that the

MD&A offers deeper insights than financial numbers alone (Audi et al., 2016; Hussainey and Walker, 2009; Wang et al., 2023) and provides a discretionary space for firms to construct corporate images (BROWN and TUCKER, 2011; Jensen and Meckling, 1994). As a result, the MD&A is an ideal domain to investigate impression management strategies and their implications for corporate accountability.

Despite substantial progress in analyzing MD&A narratives, prior research has primarily focused on qualitative attributes such as tone and readability (LOUGHRAN and MCDONALD, 2016; Bajaj et al., 2023; Tam et al., 2023). However, the link between MD&A textual narratives and Corporate Social Responsibility (CSR) outcomes remains underexplored (Ben-Amar and Belgacem, 2018; Durnev and Mangan, 2020; Wang et al., 2023). Addressing this gap, this study investigates how textual similarity within MD&A disclosures relates to CSR performance. Textual similarity refers to the extent to which the language and themes in a firm's MD&A remain consistent across consecutive reporting periods. Employing cosine similarity and Latent Semantic Indexing (LSI), we transform MD&A narratives into multi-dimensional vector representations to quantify textual similarity. High narrative similarity may signify minimal updates over time, which could either reflect stability or a deliberate attempt to obscure underperformance.



This study utilizes data from Chinese A-share-listed firms from 2010 to 2021 to examine the relationship between MD&A textual similarity and CSR performance. Our findings reveal a significant negative relationship between textual similarity in MD&A disclosures and CSR performance. These firms may rely on narrative consistency as a strategy to legitimize superficial CSR engagement, exacerbating the drawbacks of impression management. Using instrumental variables (IVs), two-stage least squares (2SLS) regression, and entropy balancing, we mitigate endogeneity and selection biases to improve the robustness of our results. The concept of textual similarity is particularly salient for stakeholders, including investors and regulators, given its implications for transparency and accountability (Peterson et al., 2015; Pradhan et al., 2015). Repeated MD&A narratives over time may indicate an effort to mask risks, limit disclosure adaptability, or reduce the informational value of corporate reports (Wang et al., 2023). For regulators, persistent homogeneity could signal noncompliance with reporting standards that mandate the disclosure of material changes (e.g., Securities and Exchange Commission (SEC), 2020). Thus, textual similarity serves as a critical dimension for evaluating whether corporate disclosures genuinely reflect evolving realities or are tools of strategic obfuscation, with significant implications for market integrity and stakeholder confidence.

This study is motivated by the need to explore how narrative consistency in MD&A disclosures reflects impression management strategies within China's regulatory and cultural environment (Hoberg and Phillips, 2010; Hoberg and Phillips, 2016; Hoberg and Phillips, 2018; Hoberg and Maksimovic, 2014). From this perspective, the research makes two key contributions. First, it introduces textual similarity over time as a new measure for evaluating CSR performance, identifying high similarity as a potential signal of passive impression management. Second, it offers an in-depth analysis of these disclosure strategies in China, a critical but distinct emerging market. The regulatory and institutional context for MD&A in China differs substantially from that in mature Western markets. In the West, disclosure practices are largely shaped by robust investor protection laws, significant litigation risk, and powerful institutional investors who demand transparency (e.g., BROWN and TUCKER, 2011). In contrast, China's capital markets operate within a framework of strong state influence, relationship-based governance, and rapidly evolving regulations, such as the intensified oversight on MD&A disclosures after 2012 (Zhang et al., 2022; Zhu and Zhang, 2024). Furthermore, the prevalence of state-owned enterprises (SOEs) creates a dual system of accountability that diverges from the purely market-driven incentives common in the West (Gao et al., 2023; Lin et al., 2024). These differences create a unique setting to investigate how firms use narrative discretion (Zhu and Zhang, 2024). Examining these dynamics within the Chinese context provides a

more nuanced understanding of impression management in non-Western markets.

2. Literature Review and Hypothesis Development

2.1 Impression Management (IM) and MD&A

Impression management (IM) involves strategic efforts to influence stakeholders' perceptions, shaping a favorable image and mitigating reputational risks. Within corporate reporting, the MD&A section serves as a key platform for IM, allowing organizations to craft narratives that bridge information gaps and provide qualitative insights into operations, strategies, and future prospects (Audi et al., 2016; Securities and Exchange Commission (SEC), 2020). While MD&A disclosures enhance stakeholders' understanding of financial health, their flexibility also enables selective emphasis of strengths, downplaying of weaknesses, and narratives of optimism, particularly during crises or underperformance (Bochkay and Levine, 2019; Muslu et al., 2015; Rim and Ferguson, 2020).

Agency theory highlights how managers exploit this discretion to shape public perception and reduce uncertainties, often incorporating sustainability-focused narratives to attract socially conscious investors (Davidson et al., 2004; Wang et al., 2023). Techniques include attributing causality for outcomes, reframing adverse events, and symbolic compliance with regulations, especially during crises (Leung et al., 2015; Rezaee et al., 2024). However, such practices can obscure inefficiencies and create a facade of accountability rather than meaningful transparency (Du and Yu, 2021; Tata and Prasad, 2015).

2.2 The Intersection of IM, CSR, and MD&A

CSR performance reflects an organization's contributions to social, environmental, and economic sustainability, shaping both corporate identity and external perceptions (Carroll, 1991; Porter and Kramer, 2006; Timbate and Park, 2018). The effectiveness of CSR initiatives depends not only on substantive actions but also on communication strategies that address societal expectations and build stakeholder trust (Liu et al., 2022; Timbate and Park, 2018). However, firms often employ IM techniques to influence perceptions of CSR performance. Such strategies include selectively emphasizing favorable aspects or highlighting visible yet non-material initiatives, which can erode accountability and stakeholder trust (Cho et al., 2024). These practices obscure true performance and limit stakeholders' ability to assess meaningful progress (Tata and Prasad, 2015; Muslu et al., 2015).

While the MD&A section is primarily designed to provide an overview of a firm's financial and operational performance, studies suggest it is also frequently used for impression management (Godfrey et al., 2003; Khan et al., 2021). In this context, firms may strategically craft MD&A narratives to portray favorable CSR performance, down-

playing failures or unmet commitments, and shifting attention away from operational weaknesses (Lu et al., 2019; Wei et al., 2024). Although CSR-specific content is not always explicitly included in the MD&A, its tone, structure, and adaptability can still provide indirect insights into managerial priorities and organizational attitudes toward CSR (Ben-Amar and Belgacem, 2018).

From the perspective of agency theory, patterns of narrative repetition, vagueness, or redundancy in MD&A disclosures are often indicative of managerial opportunism, particularly when such narratives diverge from measurable CSR outcomes (Du and Yu, 2021; Jensen and Meckling, 1994). During times of operational or financial crises, firms may rely on overly generic or symbolic language to obscure CSR inefficiencies, in effect masking weak performance with narratives that symbolize, rather than substantiate, CSR engagement (Leung et al., 2015; Yang and Liu, 2017). This behavior reflects a declining commitment to meaningful CSR practices, shifting focus toward maintaining an appearance of performance rather than achieving substantive outcomes (Xu et al., 2024).

2.3 Text Similarity and CSR Performance

Textual similarity serves as a signal of managerial intent, adaptability, and transparency (Xu et al., 2024; Kong et al., 2021). Studies reveal that repetitive and static narratives of corporate reporting often reflect limited responsiveness of firms to evolving challenges (Dailey and Browning, 2014; Dyer et al., 2017; Gao et al., 2023). High textual similarity in MD&A disclosures often reflects a lack of meaningful updates, which can signal either managerial inertia or deliberate impression management aimed at maintaining a stable corporate image (Xu et al., 2024). This issue is particularly pronounced in regulatory or investment environments with inadequate oversight, where weak monitoring fails to detect such practices (Nazari et al., 2017; Pope et al., 2023).

CSR initiatives, being inherently dynamic, require disclosures that evolve to reflect ongoing actions and shifting stakeholder expectations (Du and Yu, 2021; Huang et al., 2022). However, high textual similarity often implies superficial, symbolic CSR engagement rather than substantive efforts (Bajaj et al., 2023; Yu and Bondi, 2019; Gorovaia and Makrominas, 2025). Persistent similarity in MD&A disclosures suggests stagnation, with management prioritizing consistency and stability over meaningful updates (Hoberg and Phillips, 2010; Jensen and Meckling, 1994; Leung et al., 2015). Excessive similarity reflects impression management or narrative stagnation, which limits stakeholders' understanding of CSR performance and exacerbates trust deficits (Huang et al., 2022; Qian and Zhu, 2019).

Firms with poor CSR performance frequently employ obfuscation tactics, such as vague or overly complex language, to conceal underperformance and deflect scrutiny

(Bochkay and Levine, 2019; De Franco et al., 2020). A lack of linguistic diversity in such disclosures reflects calculated efforts to project a positive image and diminish accountability (Casonato et al., 2019; Qian and Zhu, 2019). These strategies not only mask a weak commitment to CSR initiatives but also limit stakeholders' understanding of actual progress (De Franco et al., 2020).

While corporate transparency has been widely studied, most research has focused on the content of CSR disclosures, leaving gaps in understanding the structural and linguistic characteristics of MD&A narratives—particularly textual similarity—as potential signals of managerial intent (BROWN and TUCKER, 2011; Xu et al., 2024). This study addresses this gap by hypothesizing that a lack of narrative diversity in MD&A communications reflects declining organizational commitment to CSR, likely forecasting weaker future performance. Fundamentally, such narrative stagnation degrades the quality of the firm's information environment, creating ambiguity for stakeholders attempting to assess the firm's true commitment. This leads to our first-order prediction:

Hypothesis 1: The level of similarity of MD&A text is inversely related to future CSR performance.

2.4 The Mediating Role of Information Asymmetry

While our primary hypothesis (H1) posits a negative relationship between the textual similarity of MD&A sections and a firm's future CSR performance, the theoretical channel through which this effect operates requires explicit articulation. We argue that this relationship is likely not direct but is instead mediated by the degradation of the firm's information environment, a process that ultimately fosters managerial opportunism.

Theories of impression management and signaling are central to understanding how firms strategically manage their disclosures. High textual similarity in the MD&A section, where firms reuse prior narratives, is often interpreted as a form of passive impression management designed to obscure poor performance or increase information opacity, rather than providing new, substantive information (BROWN and TUCKER, 2011; Wang et al., 2023). High textual similarity is essentially a lack of information updating and a manifestation of disclosure content inertia. Its consequences are similar to other low quality disclosure features, such as ambiguity or low readability, which collectively lead to the deterioration of the information environment (BROWN and TUCKER, 2011; Lin et al., 2024). According to signaling theory, such disclosure choices send negative signals to the capital market about the firm's underlying information environment (DIAMOND and VERRECCHIA, 1991). This obfuscation increases information asymmetry between managers and external stakeholders, including financial analysts (LARKER and ZAKOLYUKINA, 2012). Recent study confirms that repetitive narratives, whether in Environmental,

social, and governance (ESG) reports or MD&A, are associated with higher information asymmetry and stock price synchronicity (Xu et al., 2024).

When faced with a less transparent information environment, analysts find it more difficult to produce accurate and consistent earnings forecasts. A lack of substantive, firm specific disclosure forces analysts to rely more on their own assumptions and varied private information, leading to greater disagreement among them (Lang and Lundholm, 1996; Gao et al., 2023). Consequently, analyst forecast dispersion, the standard deviation of their earnings forecasts, widens. We choose analyst forecast dispersion as the core proxy for information asymmetry because it not only reflects the transparency of the information environment but also directly captures the substantial impact of information uncertainty on the judgment of key market intermediaries, namely analysts (DIAMOND and VERRECCHIA, 1991), making it one of the most direct measures of the consequences of information asymmetry. Empirical research in the Chinese context consistently demonstrates that lower quality or opaque disclosures are associated with higher forecast dispersion (Chen et al., 2014; Wu et al., 2013). Therefore, MD&A narrative stagnation is not just a repetition of content but also meta information about the quality of the firm's information environment, directly affecting the information processing costs and uncertainty of capital market intermediaries (Lang and Lundholm, 1996; Liu et al., 2022). Thus, as firms use repetitive MD&A narratives as a tool for impression management (Wei et al., 2024), we expect higher MD&A textual similarity to increase analyst forecast dispersion by exacerbating information asymmetry.

Elevated information asymmetry, as captured by high forecast dispersion, impairs the monitoring function of financial analysts and other market participants. This supervisory failure resulting from information asymmetry creates space for managerial opportunism, exacerbating second-class agency problems (Bushman and Smith, 2001). When analysts cannot confidently assess a firm's future prospects, they are less able to hold management accountable for non-financial performance. Since CSR activities are typically long-term investments with returns that are difficult to verify in the short term, managers are more motivated to shift resources from these "soft" inputs to more easily observable short-term performance metrics when external oversight is weakened. This leads to a reduction in substantive future CSR investment and a decline in performance. Therefore, analyst forecast dispersion serves as a crucial mediating channel through which the quality of narrative disclosures affects a firm's substantive CSR actions. This leads to the formulation of the second hypothesis:

Hypothesis 2: Analyst forecast dispersion mediates the negative relationship between MD&A textual similarity and future CSR performance.

3. Data and Methodology

3.1 Sample Selection

Our sample comprises data from companies listed on the Chinese A-shares market from 2010 to 2021. During the data collection process, we applied the following screening procedures. First, we excluded companies designated as ST or *ST (Hu and Zheng, 2015). Second, we removed firms in the financial and insurance industries based on the 2012 industry classification guidelines from the China Securities Regulatory Commission. Third, we eliminated observations with missing values for any of the essential variables used in our analysis. Finally, to mitigate the influence of outliers, all continuous variables were winsorized at the 1% and 99% percentiles. This data screening process resulted in a final sample of 19,155 firm-year observations.

3.2 Dependent Variable

CSR performance is quantified using the CSR rating from Hexun, a significant third-party rating system in China. Since 2010, Hexun's CSR ratings have acquired widespread recognition as a dependable proxy for gauging the CSR performance of Chinese enterprises (e.g., Yang et al., 2019; Zhong et al., 2019)ⁱ (Hexun CSR rating is a widely used evaluation system for assessing the CSR performance of Chinese firms. It evaluates companies across key dimensions, including shareholder equity responsibility, employee responsibility, supplier, customer and consumer rights responsibility, environmental responsibility, and social responsibility contribution. The Hexun framework consists of 5 level-1 indicators, 13 level-2 indicators, and 37 level-3 indicators (Zhong et al., 2019).).

3.3 Independent Variable

This study builds on the approach of Brown and Tucker (BROWN and TUCKER, 2011) to measure textual similarity in MD&A documents. The MD&A reports are sourced from the CNRDS database. Using the Jieba segmentation tool in Python, we segmented the MD&A text into smaller components for analysis. During this process, we removed unnecessary elements such as numbers, punctuation, images, and tables (Li et al., 2023). To analyze the text, we applied Latent Semantic Indexing (LSI) and a widely used method called cosine similarity. LSI converts text into a mathematical representation—a multi-dimensional vector—where each word represents a unique dimension in the space. The frequency of a word determines the size (or magnitude) of the vector in that dimension (Pradhan et al., 2015). Finally, we calculated similarity values between MD&A documents to measure how similar they are, which evaluates the consistency or repetitiveness of the language used across the MD&A documents.

3.4 The Cosine Similarity Between Two Texts (t_1, t_2)

$$Similarity(t_1, t_2) = \frac{\sum_{i=1}^n t_{1i}t_{2i}}{\sqrt{\sum_{i=1}^n t_{1i}^2} \times \sqrt{\sum_{i=1}^n t_{2i}^2}} \times 100 \quad (1)$$

Textual similarity is measured using the cosine similarity equation, which calculates the cosine of the angle between two vectors representing the content of two documents in a multi-dimensional space. Each document is mapped to an N-dimensional vector, where:

$$t_1 = (w_1, w_2, \dots, w_{n-1}, w_n)$$

$$t_2 = (q_1, q_2, \dots, q_{n-1}, q_n)$$

Here, w_i and q_i denote the frequency of each word (i) in the documents, with n being the total number of unique words. A smaller angle between the two vectors indicates a higher degree of similarity between the documents. This method quantifies how similar MD&A narratives are by analyzing the frequency patterns of words, either between firms or across time. For instance, assume we are calculating the similarity between two MD&A documents for Firm A across Year 1 (t_1) and Year 2 (t_2). The illustration of the similarity score calculation is as follows:

Word	Frequency in t_1	Frequency in t_2
“profit”	4	5
“growth”	3	3
“investment”	2	0
“debt”	1	2

3.5 Control Variables

This study includes a multitude of control variables, including company size (Size), leverage ratio (Lev), overall quality of performance (TobinQ), cash conversion rate (Cashflow), board composition (Board), proportion of independent directors (Indep), managerial shareholding (Mshare), dual leadership structure (Dual), predominant shareholder stake (Top1hold), state ownership (SOE), Network Centrality Degree (Network), and whether the firm’s financial statements are audited by one of the Big Four accounting firms (Big4). Moreover, the research incorporates textual controls for word count (WordsNO) and sentiment (Emotion) in the Management’s Discussion and Analysis (MD&A) narratives, in line with the methodology set forth by Wang et al. (2023). Detailed definitions of all vari-

ables are presented in Table 1. Data of control variables are sourced from CSMAR.

3.6 Models

To test our first hypothesis (H1) that MD&A textual similarity negatively affects a firm’s subsequent CSR performance, we construct the following baseline regression model:

$$CSRScore_{i,t+1} = \alpha_0 + \alpha_1 Similarity_{i,t} + \sum \alpha_i Controls_{i,t} + Year + Industry + \varepsilon_{i,t} \quad (2)$$

In this model, the dependent variable, $CSRScore_{i,t+1}$, is the CSR performance of firm i in the next period. The key independent variable is $Similarity_{i,t}$, which measures the textual similarity of the MD&A section for firm i in the current period. The model includes a comprehensive set of firm-level control variables ($Controls_{i,t}$) as defined in Table 1. We also incorporate Year and Industry fixed effects to control for the time-variant macroeconomic shocks and stable industry-specific characteristics. $\varepsilon_{i,t}$ represents the stochastic error term.

Furthermore, to test our second hypothesis (H2), which posits that analyst forecast dispersion mediates the negative relationship between MD&A textual similarity and future CSR performance, we employ the causal steps approach (Baron and Kenny, 1986). This requires estimating two additional models.

First, to test the effect of MD&A textual similarity on the proposed mediator (path a), we estimate Model (3), where analyst forecast dispersion ($Dispersion_{i,t}$) is the dependent variable:

$$Dispersion_{i,t} = \gamma_0 + \gamma_1 Similarity_{i,t} + \sum \gamma_i Controls_{i,t} + Year + Industry + \varepsilon_{i,t} \quad (3)$$

Second, to test the effect of the mediator on future CSR performance while controlling for MD&A textual similarity (paths b and c'), we estimate Model (4):

$$CSRScore_{i,t+1} = \delta_0 + \delta_1 Similarity_{i,t} + \delta_2 Dispersion_{i,t} + \sum \delta_i Controls_{i,t} + Year + Industry + \varepsilon_{i,t} \quad (4)$$

Cosine similarity formula components	Calculation
1. Dot Product	$((4 \times 5) + (3 \times 3) + (2 \times 0) + (1 \times 2) = 31)$
2. Magnitude of t_1 $\sqrt{\sum_{i=1}^n t_{1i}^2}$	$\sqrt{16 + 9 + 4 + 1} = \sqrt{30} = 5.477$
3. Magnitude of t_2 $\sqrt{\sum_{i=1}^n t_{2i}^2}$	$\sqrt{25 + 9 + 0 + 4} = \sqrt{38} = 6.164$
4. Cosine Similarity $Similarity(t_1, t_2)$	$\frac{DotProduct}{ t_1 \times t_2 } = \frac{31}{5.477 \times 6.164} \times 100 = 91.8$

Table 1. Variable definition.

Variable name	Variable definition
CSRScore	The Corporate Social Responsibility (CSR) performance score of the listed company
Similarity	Firm-specific measure capturing textual similarity between the Management's Discussion and Analysis (MD&A) sections of a firm's current and prior-year annual reports. Calculated using cosine similarity, this metric identifies repetitive narratives or a lack of updates in the content.
Dispersion	Analyst forecast dispersion, used as a proxy for information asymmetry. Measured as the standard deviation of analysts' one-year-ahead earnings per share (EPS) forecasts, scaled by the absolute value of the mean forecast to mitigate the influence of extreme values.
Network	Centrality within the independent directors' network, representing a firm's connection strength through shared board members. Network Centrality Degree is calculated as $Network_i = \frac{\sum_j X_{ji}}{g-1}$, where χ_{ij} represents a network connection where, if independent directors i and j serve together on at least one board, χ_{ij} is assigned a value of 1; if not, the value is 0. To mitigate the impact of board size, standardization is performed using the total number of directors (g) minus 1. A higher Degree value indicates greater centrality within the networkⁱⁱ (iiFor example, consider a firm with three independent directors: Director A, Director B, and Director C. Director A and Director B also sit together on the board of "Company X". (This connection counts as 1). Director A and Director C do not share any other board seats. (This counts as 0). Director B and Director C also sit together on the board of "Company Y". (This connection counts as 1). The sum of connections is 1 + 0 + 1 = 2. The total number of independent directors (g) is 3. The Network Centrality score would be calculated as: $2/(3 - 1) = 1.0$. This score reflects how interconnected this firm's board is.).
Size	The natural logarithm of total assets at the end of the year
Lev	Leverage ratio, calculated as total year-end liabilities divided by total year-end assets.
Cashflow	Cash flow performance, measured as net cash flow from operating activities divided by total assets at the period-end.
TobinQ	Defined as the ratio of market value (A) to total assets. The market value (A) is calculated as (A-Share Quantity × A-Share Closing Price) + (B-Share Quantity × B-Share Closing Price) + [(Total Shares – A Share – B Share) × (Total Owners' Equity at End of Period/Ending Paid-in Capital)] + Total Liabilities at End of Current Period.
Board	Total number of directors on the company's board
Independ	Ratio of independent directors to the total number of directors
Mshare	Management shareholding ratio, calculated as the proportion of shares held by directors, supervisors, and senior executives to the total share capital.
Dual	Dummy variable equal to 1 if the chairman and general manager (CEO) are the same person, indicating CEO/Chair duality; equal to 0 otherwise.
Top1hold	Ownership concentration, measured as the ratio of shares held by the largest shareholder to the total number of shares in the company.
SOE	Dummy variable equal to 1 for state-owned enterprises (SOEs) and 0 for non-state-owned enterprises. An enterprise is classified as state-owned if its ultimate controlling entity includes a state-owned enterprise, administrative institution, public service unit, central government institution, or local government institution. If multiple controlling entities exist, the company is still classified as state-owned if at least one controlling entity qualifies as an SOE.
Big4	Dummy variable equal to 1 if the company's auditor is one of the Big Four accounting firms (Deloitte, PwC, Ernst & Young, KPMG); equal to 0 otherwise.
WordsNO	Total word count of the MD&A section
Emotion	Sentiment ratio, calculated as the ratio of negative words to positive words in the MD&A section

Here, *Dispersion* is calculated as follows:
 $Dispersion_{i,t} = \frac{Std(FEPS_{i,t})}{[Mean(FEPS_{i,t})]}$ Where $FEPS_{i,t}$ represents the set of all available one-year-ahead EPS forecasts for firm i in year t .

3.7 $Std(FEPS_{i,t})$ is the Standard Deviation for These Forecasts and $Mean(FEPS_{i,t})$ is Their Arithmetic Mean

A significant coefficient for γ_1 in Model (3) and for δ_2 in Model (4) would provide initial support for the mediation mechanism. To formally test the significance of the indirect effect (the product of paths a and b), and in line with modern best practices that address the statistical limitations of the Sobel test, we will use a Bootstrap procedure with 5000 replications to estimate the confidence interval of the indirect effect. A confidence interval that does not contain zero would provide robust evidence for the mediating role of analyst forecast dispersion.

4. Discussion and Analysis

4.1 Descriptive Statistics and Correlation Analysis

Table 2 describes the descriptive statistics for our analysis. The average similarity score of the MD&A disclosures among the sampled Chinese companies stands at 93.548, which indicates a high degree of homogeneity in how these firms communicate within their MD&A sections (Shen and Zhou, 2024; Xue et al., 2023). However, this uniformity is juxtaposed with pronounced individual variations in scores, ranging markedly from 99.150 at the upper limit to 60.780 at the lower, thereby signifying substantial divergence in the textual similarity of disclosures among different firms. In terms of CSR performance, the average score within our sample is determined to be 24.543, denoting considerable challenges faced by sample companies in fulfilling their CSR obligations (Fu et al., 2023). Additionally, the CSR scores demonstrate a wide distribution, extending from a low of 0.010 to a high of 79.350. Furthermore, the analyst forecast dispersion (Dispersion), our proxy for information asymmetry, has a mean of 0.028 but a wide range from 0.000 to 0.893. This significant variability, highlighted by a standard deviation of 0.045, suggests that while most firms maintain a clear information environment, a subset of firms generates substantial analyst disagreement, consistent with the use of narrative obfuscation as an impression management tactic.

The findings depicted in Table 3 detail the Pearson correlations among the variables. Notably, the table provides strong preliminary support for our core hypotheses. First, it reveals a statistically significant negative correlation coefficient (-0.109) between future CSR performance (CSRScore) and the similarity of MD&A texts (Similarity) at the 1% significance level, offering initial validation for H1. More importantly, the correlations provide descriptive evidence for the mediating role of analyst forecast dispersion (Dispersion), as proposed in H2. There is a significant positive correlation between Similarity and Dispersion

Table 2. Descriptive statistics.

VarName	Obs	Mean	SD	Min	Max
CSRScore	19,155	24.543	14.795	0.010	79.350
Similarity	19,155	93.548	4.637	60.780	99.150
Dispersion	19,155	0.028	0.045	0.000	0.893
Network	19,155	0.392	0.358	0.000	2.000
Size	19,155	22.154	1.256	19.692	26.360
Lev	19,155	0.416	0.202	0.032	0.901
Cashflow	19,155	0.048	0.068	-0.196	0.259
TobinQ	19,155	2.042	1.315	0.785	15.401
Board	19,155	2.138	0.194	1.609	2.708
Independ	19,155	37.407	5.328	28.570	60.000
Mshare	19,155	13.485	19.929	0.000	70.381
Dual	19,155	0.268	0.443	0.000	1.000
Top1hold	19,155	37.047	14.826	6.870	76.080
SOE	19,155	0.378	0.485	0.000	1.000
Big4	19,155	0.943	0.232	0.000	1.000
WordsNO	19,155	9.279	0.449	2.197	11.706
Emotion	19,155	0.575	0.154	0.201	1.558

Obs, Observations; SD, Standard Deviation.

(0.125, $p < 0.01$), and a significant negative correlation between Dispersion and future CSRScore (-0.081 , $p < 0.01$). This pattern suggests that higher MD&A textual similarity is associated with greater information asymmetry, which in turn is linked to poorer subsequent CSR performance. The study further addresses the issue of multicollinearity among the independent variables within the linear models. By implementing a Variance Inflation Factor (VIF) test, the evidence demonstrates that all VIF values sit comfortably below the conventional threshold of 5, with the mean value falling below 2. This assessment attests to the absence of significant multicollinearity within the model variables.

4.2 Effects of MD&A Similarity on Subsequent CSR Performance

Table 4 presents the main regression results for our primary research hypothesis (H1), which posits that higher MD&A textual similarity is associated with poorer subsequent CSR performance. To ensure the robustness of our findings and in direct response to the editor's feedback, we present the results using a two-tiered fixed-effects strategy, with the baseline model presented in Column (1) and a more stringent robustness check in Column (2).

In Column (1), we report the results of our baseline model, which includes the full set of control variables and employs Year and Industry fixed effects. The coefficient for our key independent variable, Similarity, is -0.050 and is statistically significant at the 1% level. This result provides strong initial support for H1, indicating that, on average, firms with more repetitive MD&A narratives tend to exhibit worse CSR performance in the following year.

In Column (2), we present a more rigorous test by replacing industry fixed effects with firm fixed effects. This

Table 3. Pearson correlation matrix.

	CSRScore	Similarity	Dispersion	Network	Size	Lev	Cashflow	TobinQ	Board	Independ	Mshare	Dual	Top1hold	SOE	Big4	WordsNO	Emotion
CSRScore	1																
Similarity	-0.109***	1															
Dispersion	-0.081***	0.125***	1														
Network	-0.018**	0.057***	0.041*	1													
Size	0.213***	-0.058***	-0.095*	0.062***	1												
Lev	0.022***	-0.107***	0.063*	-0.014**	0.538***	1											
Cashflow	0.174***	0.020***	-0.033*	0.025***	0.055***	-0.155***	1										
TobinQ	-0.002	-0.042***	0.051*	-0.036***	-0.391***	-0.268***	0.084***	1									
Board	0.117***	-0.054***	0.015	-0.057***	0.252***	0.164***	0.036***	-0.124***	1								
Independ	-0.007	0.015**	-0.021	0.026***	0.011	-0.018**	-0.006	0.039***	-0.530***	1							
Mshare	-0.074***	0.140***	0.088*	0.047***	-0.360***	-0.329***	-0.011	0.022**	-0.209***	0.078***	1						
Dual	-0.068***	0.066***	0.045*	0.032***	-0.185***	-0.137***	-0.013*	0.062***	-0.190***	0.114***	0.262***	1					
Top1hold	0.128***	-0.019***	-0.052*	0.008	0.108***	-0.014*	0.102***	-0.090***	-0.023**	0.057***	0.070***	-0.017**	1				
SOE	0.143***	-0.111***	-0.077*	-0.062***	0.362***	0.306***	-0.004	-0.132***	0.285***	-0.059***	-0.504***	-0.296***	0.097***	1			
Big4	-0.153***	0.014**	-0.065*	-0.025***	-0.350***	-0.128***	-0.079***	0.083***	-0.081***	-0.034***	0.129***	0.069***	-0.111***	-0.143***	1		
WordsNO	0.040***	0.211***	0.031*	0.048***	0.036***	-0.074***	-0.059***	-0.090***	-0.012	0.014**	0.192**	0.086**	0.015**	-0.153**	0.015**	1	
Emotion	-0.126***	-0.088***	0.059*	0.003	0.002	0.035***	-0.007	0.014*	-0.017**	-0.006	-0.119***	-0.028***	-0.035***	0.015**	-0.016**	-0.230***	1

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

specification controls for all time-invariant unobserved firm characteristics (e.g., corporate culture, managerial style), thus mitigating concerns about omitted variable bias. The coefficient for Similarity remains negative and statistically significant (-0.048 , $p < 0.01$). The stability of our main finding, even after absorbing all firm-specific heterogeneity, strengthens the causal interpretation of our results.

Table 4. The impact of MD&A textual similarity on future CSR performance.

Variables	(1)	(2)
	CSRScore	CSRScore
Similarity	-0.050^{***} (0.024)	-0.048^{***} (0.023)
Control Variables	Included	Included
Year FE	YES	YES
Industry FE	YES	NO
Firm FE	NO	YES
Observations	19,155	19,155
Adj. R-squared	0.473	0.512

Notes: This table presents the Ordinary Least Squares (OLS) Regression results for the effect of MD&A textual similarity on future CSR performance. The dependent variable is CSRScore measured in year $t+1$. The key independent variable is Similarity measured in year t . Standard errors, clustered at the firm level, are reported in parentheses. Control variables include Size, Lev, Cashflow, TobinQ, Board, Independ, Mshare, Dual, Top1hold, SOE, Big4, WordsNO, and Emotion. Standard errors, are in parentheses. Robust standard errors are in parentheses, *** $p < 0.01$. FE, Fixed Effects.

These findings suggest that excessive textual similarity in MD&A disclosures may reflect managerial inertia, signaling superficial engagement with CSR rather than genuine commitment. High similarity could also indicate efforts to obscure unfavorable financial or operational realities by reducing MD&A disclosures to mere formalities (Wang et al., 2023). This behavior aligns with passive impression management, allowing firms to create an illusion of stability or strategic direction without actively engaging in meaningful CSR activities (Im et al., 2021). Unlike active impression management tactics, such as conspicuous CSR actions that may risk being perceived as 'greenwashing' (Bansal and Clelland, 2004), these passive strategies undermine transparency and accountability. This lack of accountability leads to poorer substantive CSR outcomes.

4.3 Mediation Analysis of Analyst Forecast Dispersion

To test our second hypothesis (H2), we examine whether information asymmetry mediates the negative relationship between MD&A textual similarity and future CSR performance. We employ the causal steps approach outlined by Baron and Kenny (1986) and use analyst forecast dispersion (Dispersion) as a proxy for information asymmetry. Table 5 presents the results.

Table 5. The mediating role of analyst forecast dispersion.

Variables	(1)	(2)	(3)
	CSRScore	Dispersion	CSRScore
Similarity	-0.048^{***} (0.018)	0.009^{***} (0.003)	-0.035^{**} (0.017)
Dispersion			-1.450^{***} (0.481)
Control Variables	Included	Included	Included
Year FE & Industry FE	YES	YES	YES
Observations	19,155	19,155	19,155
Adj. R-squared	0.473	0.215	0.478
Sobel Test (z-value)			-4.18^{***}
Bootstrap 95% CI			$[-0.019, -0.008]$

Notes: This table presents the results of the mediation analysis testing the role of Analyst Forecast Dispersion (Dispersion) in the relationship between MD&A textual similarity (Similarity) and future CSR performance (CSRScore). Column (1) reports the total effect (Path c). Column (2) reports the effect of the independent variable on the mediator (Path a). Column (3) reports the effects of the independent and mediating variables on the dependent variable (Paths c' and b). Robust standard errors are in parentheses. ** $p < 0.05$, *** $p < 0.01$.

Column (1) repeats our baseline finding from Table 4. This confirms the significant total effect of Similarity on future CSRScore. The coefficient is -0.048 and is statistically significant at the 1% level, satisfying the first condition for mediation.

Column (2) tests the effect of our independent variable on the mediator. The results show that Similarity has a positive and highly significant coefficient on Dispersion ($\beta = 0.009$, $p < 0.01$). This finding supports the first part of our proposed mechanism. Firms that use more repetitive MD&A narratives effectively degrade their information environment, leading to greater disagreement and uncertainty among financial analysts.

Column (3) presents the results of the full model, where both Similarity and Dispersion predict CSRScore. Two key insights emerge. First, the coefficient on Dispersion is negative and highly significant ($\beta = -1.450$, $p < 0.01$). This confirms that increased information asymmetry is directly associated with poorer subsequent CSR performance. The result holds even after controlling for MD&A similarity. This supports the second part of our mechanism, where a weakened information environment allows for managerial opportunism that negatively impacts CSR engagement.

Second, when the mediator (Dispersion) is included, the coefficient on Similarity remains statistically significant but is notably reduced in magnitude from -0.048 to -0.035 . The drop in the coefficient size indicates that Dispersion partially mediates the relationship between Similarity and CSRScore.

Table 6. Heterogeneity analysis.

	(1)	(2)	(3)	(4)	(5)	(6)
Sub-sample:	Non-SOE	SOE	Eastern	Non-Eastern	High Attention	Low Attention
Variables	CSRScore	CSRScore	CSRScore	CSRScore	CSRScore	CSRScore
Similarity	−0.067*** (0.027)	−0.020 (0.045)	−0.058** (0.028)	−0.024 (0.048)	−0.063* (0.034)	−0.008 (0.039)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Industry & Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Chow Test	Significant		Significant		Significant	
Observations	11,910	7245	13,831	5307	11,448	7690
Adj. R-squared	0.139	0.270	0.205	0.186	0.228	0.146

Notes: This table presents OLS regression results for the heterogeneity analysis. The dependent variable, CSRScore, is the corporate social responsibility score from the Hexun rating system. The key independent variable, Similarity, is the cosine similarity of a firm's MD&A text between the current and prior year. The sample is split into sub-samples based on: (1) Ownership: State-Owned Enterprises (SOEs) vs. Non-State-Owned Enterprises (Non-SOEs) in Columns (1) and (2); (2) Location: firms headquartered in China's eastern provinces vs. those in central and western (Non-Eastern) provinces in Columns (3) and (4); and (3) Analyst Attention: firms with above-median analyst coverage (High Attention) vs. those with below-median coverage (Low Attention) in Columns (5) and (6). All models include the same set of control variables as in the main analysis (Table 4). Standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

To formally test the significance of this indirect effect, we performed both the Sobel test and a bootstrap analysis. The Sobel test yields a z-value of -4.18 , significant at the 1% level. More importantly, the bootstrap analysis with 5000 replications produces a 95% confidence interval for the indirect effect of $[-0.019, -0.008]$. This interval does not contain zero, so we can confidently conclude the mediating effect is statistically significant.

These results provide robust empirical evidence for H2. The findings demonstrate that firms engaging in passive impression management through repetitive MD&A narratives foster an opaque information environment (BROWN and TUCKER, 2011; Wang et al., 2023). This elevated information asymmetry is captured by a significant increase in analyst forecast dispersion. Such opacity impairs the external monitoring function of the market, particularly that of financial analysts (Lang and Lundholm, 1996; Xu et al., 2024). Within this environment of weakened oversight, managers have greater latitude for opportunistic behavior, leading them to underinvest in long-term, hard-to-verify activities like corporate social responsibility (Gao et al., 2023). Therefore, our analysis confirms that analyst forecast dispersion serves as a significant mediating channel through which MD&A textual similarity negatively impacts subsequent CSR performance.

4.4 Heterogeneity Analysis

To explore the contextual boundaries of our main finding, we conduct several heterogeneity analyses based on the baseline model specification, which includes Industry and Year Fixed Effects as stipulated by the reviewers.

4.5 Heterogeneity Across Corporate Ownership Structure

Corporate ownership structures play a critical role in shaping management practices, including CSR engagement and MD&A disclosure strategies (Bova and Yang, 2018). We examine the relationship between MD&A textual similarity and CSR performance across ownership types.

Table 6, Columns (1) and (2), demonstrates a significant negative correlation for non-SOEs (coefficient = -0.067 , $p < 0.01$), while SOEs show an insignificant correlation (coefficient = -0.020). These findings highlight the distinct strategies employed by non-SOEs and SOEs (Kao et al., 2018). For non-SOEs, increased MD&A textual similarity may signify strategic neglect in CSR practices. In contrast, tighter government oversight over SOEs may mitigate the negative impact of disclosure repetition (Gao, 2011).

4.6 Heterogeneity Across Corporate Headquarter Locations

Geographical diversity significantly influences corporate behavior, as socio-economic and regulatory factors vary across regions (Derouiche et al., 2016). Table 6, Columns (3) and (4), shows differing correlations between MD&A textual similarity and CSR performance in China's economically diverse regions.

In the economically dynamic eastern provinces, the negative relationship is more pronounced and statistically significant (coefficient = -0.058 , $p < 0.05$). Conversely, firms in the comparatively underdeveloped central-western regions show a weaker, insignificant relationship (coefficient = -0.024). This suggests that firms in more developed regions face greater market pressure for high-quality disclosures and proactive CSR initiatives (Husted et al., 2016).

4.7 Heterogeneity Across Analyst Attention

Analyst attention substantially impacts the perception of corporate information. Firms with significant analyst coverage face greater scrutiny, influencing corporate behavior and reputation (Hirshleifer and Teoh, 2003).

Table 6, Columns (5) and (6), divide the sample into firms with high and low analyst attention. For firms under high analyst attention, high MD&A textual similarity has a negative and marginally significant impact on CSR performance (coefficient = -0.063 , $p < 0.1$), suggesting it raises stakeholder skepticism (Hu et al., 2021). Conversely, for firms with low analyst attention, similar textual disclosures have a negligible and insignificant impact (coefficient = -0.008).

4.8 Endogeneity Tests

To address potential endogeneity concerns, we employ three distinct econometric strategies: a dynamic panel GMM estimation to control for persistence and dynamic endogeneity; a two-stage least squares (2SLS) instrumental variable (IV) approach to address reverse causality and omitted variables; and the entropy balancing method to mitigate potential selection bias.

4.9 Dynamic GMM Estimation

First, to address concerns that CSR performance is persistent over time and to mitigate dynamic endogeneity, we use a two-step system GMM estimator. This model includes a lagged dependent variable to explicitly control for the persistence in a firm's CSR activities.

The results are presented in Table 7. The coefficient on the lagged CSR performance (L.CSRScore) is positive and highly significant, confirming that CSR performance is indeed persistent. More importantly, even after controlling for this dynamic effect, the coefficient on Similarity remains negative and statistically significant (-0.041 , $p < 0.05$). The model's validity is supported by the Hansen test ($p = 0.152$) and the AR(2) test ($p = 0.234$), which confirm the validity of the internal instruments and the absence of second-order serial correlation, respectively.

4.10 IV-2SLS Estimation

To address potential endogeneity arising from omitted variables or reverse causality, we employ an IV regression using a 2SLS estimator. Following prior literature (e.g., Durnev and Mangen, 2020), we use the industry-average MD&A textual similarity (excluding the focal firm) as an instrument. The rationale for this instrument is that firms within the same industry face similar competitive pressures and institutional logics, leading to comparable disclosure practices (Luo, 2003; Rugman and Verbeke, 1998).

The diagnostic tests confirm the instrument's validity. As shown in Table 8, the first-stage Sanderson-Windmeijer (S-W) F-statistic of 29.30 confirms the instrument is relevant and not weak. The validity of our IV approach also

Table 7. Dynamic gmm estimation.

Variables	(1)
	CSRScore
L.CSRScore	0.512*** (0.045)
Similarity	-0.041** (0.019)
Controls	Yes
Industry & Year FE	Yes
Observations	17,240
Hansen test (p -value)	0.152
AR(2) test (p -value)	0.234

Notes: This table presents the results from the two-step system GMM estimation. The dependent variable is CSRScore. L.CSRScore is the one-year lag of the dependent variable. Similarity is the cosine similarity of the MD&A text. All models include the full set of control variables from the main analysis. Robust standard errors are in parentheses. ** $p < 0.05$, *** $p < 0.01$. The p -values for the Hansen test of over-identifying restrictions and the Arellano-Bond test for AR(2) in first differences are reported.

depends on the untestable exclusion restriction. Crucially, the second-stage results show that the instrumented Similarity has a significant negative effect on CSRScore (coefficient = -1.405 , $p < 0.05$), providing strong support for our hypothesis.

Table 8. Instrumental variable (IV-2SLS) regression.

Model	(1)	(2)
	First Stage	Second Stage
Dependent Var:	Similarity	CSRScore
Similarity		-1.405** (0.672)
IV	0.325*** (0.064)	
Controls	Yes	Yes
Industry & Year FE	Yes	Yes
Observations	19,147	19,147
Adj. R-squared	0.278	0.103
S-W F-statistic	29.30	

Notes: This table presents the IV-2SLS results. The instrument (IV) is the industry-average Similarity. Column (1) reports the first stage, and Column (2) reports the second stage. Robust standard errors are in parentheses. ** $p < 0.05$, *** $p < 0.01$.

4.11 Entropy Balancing Method

Finally, to mitigate potential selection bias between firms with high versus low disclosure similarity, we use the entropy balancing method. This approach re-weights the

control group (low similarity) to perfectly match the covariate moments of the treatment group (high similarity). Table 9 reports the OLS regression results on this re-weighted sample. The coefficient for Similarity is -0.059 and remains significant ($p < 0.05$), indicating that our findings hold after correcting for observable differences between the groups.

Table 9. Entropy balancing method.

Variables	(1)
	CSRScore
Similarity	-0.059^{**} (0.027)
Controls	Yes
Industry & Year FE	Yes
Observations	19,155
Adj. R-squared	0.474

Notes: This table reports the OLS regression results after applying entropy balancing to the sample based on the median value of Similarity. Robust standard errors are in parentheses. $** p < 0.05$.

4.12 Robustness Tests

We conduct several additional tests to confirm the robustness of our main findings. These tests include using alternative variable measurements and different sample periods. The results are presented in Table 10.

4.13 Substitution of the Dependent Variable Measurement

We first refine the CSR performance metric to account for industry-wide trends (Clarkson et al., 2008). A new dependent variable, Adjusted CSR Performance (AdjCSR), is created by subtracting the industry-year average from each firm's raw CSR score. Column (1) of Table 10 shows that Similarity has a significant negative coefficient of -0.040 ($p < 0.05$) on this adjusted metric, reinforcing our primary hypothesis.

4.14 Substitution of Control Variables

Corporate governance structures are known to influence CSR outcomes (Velte, 2020). We test an alternative specification in our model by replacing Board size with the total number of managers (Mnumber) as a proxy for executive leadership. The result in Column (2) of Table 10 confirms a significant inverse relationship between Similarity and CSRScore (coefficient = -0.050 , $p < 0.05$). Next, we substitute Tobin's Q with Return on Assets (ROA) as the firm performance metric. This change addresses the potential endogeneity of Tobin's Q to managerial choices (Dybvig and Warachka, 2011) and uses ROA as a direct indicator of financial health (Martínez et al., 2007; Alghifari et al., 2013). As shown in Column (3), the coefficient for Similarity remains negative and significant (coefficient = -0.040 , $p < 0.05$), which further affirms our hypothesis.

4.15 Adjusting for Different Sampling Period

Regulatory enhancements in China intensified oversight on MD&A disclosures after 2012 (Liu et al., 2022). We test the robustness of our findings in this stricter regulatory environment by narrowing the sample to the 2013–2021 period. The result in Column (4) of Table 10 reveals a significant negative coefficient of -0.058 ($p < 0.05$) for Similarity. To neutralize any potential distortive effects from the COVID-19 pandemic, we also conduct a test using only pre-pandemic data from 2010–2019. The result in Column (5) shows a consistent and significant negative relationship between Similarity and CSRScore (coefficient = -0.051 , $p < 0.05$), which enhances the reliability of our findings.

5. Concluding Remarks

This study investigates the strategic use of narrative similarity in MD&A disclosures as a form of passive impression management. Our empirical results demonstrate a

Table 10. Robustness tests.

Variables	(1)	(2)	(3)	(4)	(5)
	AdjCSR	CSRScore	CSRScore	CSRScore	CSRScore
Similarity	-0.040^{**} (0.024)	-0.050^{**} (0.024)	-0.040^{**} (0.024)	-0.058^{**} (0.025)	-0.051^{**} (0.024)
Controls	Yes	Yes	Yes	Yes	Yes
Industry & Year FE	Yes	Yes	Yes	Yes	Yes
Observations	19,155	19,155	19,154	14,972	18,914
Adj. R-squared	0.385	0.475	0.478	0.406	0.476

Notes: This table presents results from several robustness tests. Column (1) uses the industry-adjusted AdjCSR as the dependent variable. Column (2) replaces Board size with the number of managers (Mnumber). Column (3) replaces TobinQ with Return on Assets (ROA). Column (4) restricts the sample to the 2013–2021 period. Column (5) restricts the sample to the pre-COVID 2010–2019 period. All models include the full set of control variables (unless substituted) and include Industry and Year Fixed Effects. Robust standard errors are in parentheses. $** p < 0.05$.

robust negative relationship between high textual similarity in MD&A reports and a firm's subsequent CSR performance. This finding, validated through a comprehensive set of endogeneity and robustness tests including dynamic GMM and IV regressions, suggests that repetitive, boilerplate disclosures often signal a lack of substantive engagement with CSR. By empirically linking a specific textual attribute to tangible performance outcomes, our research offers a novel, quantifiable measure of passive impression management and its consequences.

Our findings yield several practical and policy implications. For regulators, this study suggests that textual analysis can be a powerful supervisory tool; unusually high MD&A similarity could serve as a data-driven red flag to identify firms warranting closer scrutiny of their substantive CSR activities. For investors and analysts, MD&A textual similarity offers a simple heuristic to help unmask sophisticated "greenwashing" by flagging a potential disconnect between a firm's stated CSR goals and its actual commitment. For corporate managers, our results serve as a cautionary note that relying on boilerplate disclosures is a detectable strategy that can degrade the information environment and damage stakeholder trust, underscoring the need for authentic and dynamic reporting.

This study opens several avenues for future research. Scholars could investigate other linguistic features, such as thematic shifts or the use of forward-looking language, as predictors of CSR engagement. Examining our findings in different institutional and cultural contexts would also be a valuable extension. For instance, future research could explore how factors like a firm's social network position or the presence of specific governance mechanisms moderate the relationship we document. Such studies would continue to deepen our understanding of the complex interplay between corporate disclosure strategies and sustainable value creation in a globalized economy.

Availability of Data and Materials

The data that support the findings of this study are available from the CSMAR (China Stock Market & Accounting Research) database. Restrictions may apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data may be available from the authors upon request.

Author Contributions

RK and RZ conceived and designed the study. RK and RZ contributed to the methodology and research design. YZ made substantial contributions to [acquisition of data (e.g., defining data sources, sample screening rules, and key variable construction) and/or analysis and interpretation of the data/results]. RK conducted the empirical analyses. RK drafted the manuscript. RZ and YZ critically reviewed and revised the manuscript for important intellectual content. All authors approved the final version to be

published and agree to be accountable for all aspects of the work.

Acknowledgment

The authors would like to express their sincere gratitude to the anonymous reviewers for their insightful and constructive comments, which have significantly improved the quality of this manuscript. All remaining errors are our own.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflicts of interest.

Declaration of AI and AI-Assisted Technologies in the Writing Process

The authors declare that no artificial intelligence (AI) or AI-assisted technologies were utilized in any part of the ideation, data analysis, or writing process of this manuscript. The entire intellectual content and composition of this work are solely the product of the listed human authors.

References

- Alghifari ES, Triharjono S, Juhaeni YS. Effect of return on assets (ROA) against Tobin's Q: Studies in food and beverage company in Indonesia stock exchange years 2007-2011. *International Journal of Science and Research (IJSR)*. 2013; 2: 722–725. <https://www.ijsr.net/archive/v2i1/IJSRON2013372.pdf>
- Audi R, Loughran T, McDonald B. Trust, but Verify: MD&A Language and the Role of Trust in Corporate Culture. *Journal of Business Ethics*. 2016; 139: 551–561. <https://doi.org/10.1007/s10551-015-2659-4>
- Bajaj A, Leonard LNK, Sun L, Xing Z. Corporate social responsibility and annual report reading difficulty. *Review of Quantitative Finance and Accounting*. 2023; 60: 1393–1428. <https://doi.org/10.1007/s11156-023-01132-9>
- Bansal P, Clelland I. TALKING TRASH: LEGITIMACY, IMPRESSION MANAGEMENT, AND UNSYSTEMATIC RISK IN THE CONTEXT OF THE NATURAL ENVIRONMENT. *Academy of Management Journal*. 2004; 47: 93–103. <https://doi.org/10.2307/20159562>
- Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*. 1986; 51: 1173–1182. <https://doi.org/10.1037/0022-3514.51.6.1173>
- Ben-Amar W, Belgacem I. Do socially responsible firms provide more readable disclosures in annual reports? *Corporate Social Responsibility and Environmental Management*. 2018; 25: 1009–1018. <https://doi.org/10.1002/csr.1517>
- Bochkay K, Levine CB. Using MD&A to Improve Earnings

- Forecasts. *Journal of Accounting, Auditing & Finance*. 2019; 34: 458–482. <https://doi.org/10.1177/0148558x17722919>
- Bova F, Yang L. State-Owned Enterprises, Competition, and Disclosure. *Contemporary Accounting Research*. 2018; 35: 596–621. <https://doi.org/10.1111/1911-3846.12412>
- BROWN SV, TUCKER JW. Large-Sample Evidence on Firms' Year-over-Year MD&A Modifications. *Journal of Accounting Research*. 2011; 49: 309–346. <https://doi.org/10.1111/j.1475-679x.2010.00396.x>
- Bushman RM, Smith AJ. Financial accounting information and corporate governance. *Journal of Accounting and Economics*. 2001; 32: 237–333. [https://doi.org/10.1016/S0165-4101\(01\)00027-1](https://doi.org/10.1016/S0165-4101(01)00027-1)
- Carroll AB. The pyramid of corporate social responsibility: Toward the moral management of organizational stakeholders. *Business Horizons*. 1991; 34: 39–48. [https://doi.org/10.1016/0007-6813\(91\)90005-g](https://doi.org/10.1016/0007-6813(91)90005-g)
- Casonato F, Farneti F, Dumay J. Social capital and integrated reporting: Losing legitimacy when reporting talk is not supported by actions. *Journal of Intellectual Capital*. 2019; 20: 144–164. <https://doi.org/10.1108/jic-08-2018-0132>
- Chen Y, Wang Y, Lin L. Independent directors' board networks and controlling shareholders' tunneling behavior. *China Journal of Accounting Research*. 2014; 7: 101–118. <https://doi.org/10.1016/j.cjar.2013.09.002>
- Cho M, Krishnan GV, Cho H. Can We Trust the Trust Words in 10-Ks? *Journal of Business Ethics*. 2024; 190: 975–992. <https://doi.org/10.1007/s10551-023-05350-y>
- Clarkson PM, Li Y, Richardson GD, Vasvari FP. Revisiting the relation between environmental performance and environmental disclosure: An empirical analysis. *Accounting, Organizations and Society*. 2008; 33: 303–327. <https://doi.org/10.1016/j.aos.2007.05.003>
- Dailey SL, Browning L. Retelling Stories in Organizations: Understanding the Functions of Narrative Repetition. *Academy of Management Review*. 2014; 39: 22–43. <https://doi.org/10.5465/amr.2011.0329>
- Davidson WN, Jiraporn P, Kim YS, Nemec C. EARNINGS MANAGEMENT FOLLOWING DUALITY-CREATING SUCCESSIONS: ETHNOSTATISTICS, IMPRESSION MANAGEMENT, AND AGENCY THEORY. *Academy of Management Journal*. 2004; 47: 267–275. <https://doi.org/10.2307/20159577>
- De Franco G, Fogel-Yaari H, Li H. MD&A Textual Similarity and Auditors. *Auditing: a Journal of Practice & Theory*. 2020; 39: 105–131. <https://doi.org/10.2308/ajpt-18-016>
- Deegan C. Introduction: The legitimising effect of social and environmental disclosures—a theoretical foundation. *Accounting, Auditing & Accountability Journal*. 2002; 15: 282–311. <https://doi.org/10.1108/09513570210435852>
- Derouiche I, Jaafar K, Zemzem A. Firm geographic location and voluntary disclosure. *Journal of Multinational Financial Management*. 2016; 37: 29–47. <https://doi.org/10.1016/j.mulfin.2016.08.003>
- DIAMOND DW, VERRECCHIA RE. Disclosure, Liquidity, and the Cost of Capital. *The Journal of Finance*. 1991; 46: 1325–1359. <https://doi.org/10.1111/j.1540-6261.1991.tb04620.x>
- Du S, Yu K. Do Corporate Social Responsibility Reports Convey Value Relevant Information? Evidence from Report Readability and Tone. *Journal of Business Ethics*. 2021; 172: 253–274. <https://doi.org/10.1007/s10551-020-04496-3>
- Durnev A, Mangen C. The spillover effects of MD&A disclosures for real investment: The role of industry competition. *Journal of Accounting and Economics*. 2020; 70: 101299. <https://doi.org/10.1016/j.jacceco.2020.101299>
- Dybvig PH, Warachka M. Tobin's Q Does Not Measure Performance: Theory, Empirics, and Alternative Measures. *SSRN Electronic Journal*. 2011. <https://doi.org/10.2139/ssrn.1562444>
- Dyer TA, Lang MH, Stice-Lawrence L. Discussion of: 10-K Disclosure Repetition and Managerial Reporting Incentives: What Have We Learned and Where Do We Go with Textual Research? *Journal of Financial Reporting*. 2017; 2: 133–138. <https://doi.org/10.2308/jfir-51950>
- Fu L, Pan L, Zhao J. Can Passive Investors Improve Corporate Social Responsibility? Evidence from Chinese Listed Firms. *Emerging Markets Finance and Trade*. 2023; 59: 404–419. <https://doi.org/10.1080/1540496x.2022.2103404>
- Gao J, Hartmann FGH, Zhang M, Chen Y. The impact of CSR performance and CSR disclosure readability on investors' earnings estimates. *Accounting & Finance*. 2023; 63: 1157–1186. <https://doi.org/10.1111/acfi.12938>
- Gao Y. CSR in an emerging country: a content analysis of CSR reports of listed companies. *Baltic Journal of Management*. 2011; 6: 263–291. <https://doi.org/10.1108/17465261111131848>
- Godfrey J, Mather P, Ramsay A. Earnings and Impression Management in Financial Reports: The Case of CEO Changes. *Abacus*. 2003; 39: 95–123. <https://doi.org/10.1111/1467-6281.00122>
- Gorovaia N, Makrominas M. Identifying greenwashing in corporate-social responsibility reports using natural-language processing. *European Financial Management*. 2025; 31: 427–462. <https://doi.org/10.1111/eufm.12509>
- Highhouse S, Brooks ME, Gregarus G. An Organizational Impression Management Perspective on the Formation of Corporate Reputations. *Journal of Management*. 2009; 35: 1481–1493. <https://doi.org/10.1177/0149206309348788>
- Hirshleifer D, Teoh SH. Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics*. 2003; 36: 337–386. <https://doi.org/10.1016/j.jacceco.2003.10.002>
- Hoberg G, Maksimovic V. Redefining Financial Constraints: A Text-Based Analysis. *The Review of Financial Studies*. 2014; 28: 1312–1352. <https://doi.org/10.1093/rfs/hhu089>
- Hoberg G, Phillips G. Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis. *Review of Financial Studies*. 2010; 23: 3773–3811. <https://doi.org/10.1093/rfs/hhu089>

[//doi.org/10.1093/rfs/hhq053](https://doi.org/10.1093/rfs/hhq053)

- Hoberg G, Phillips G. Text-Based Network Industries and Endogenous Product Differentiation. *Journal of Political Economy*. 2016; 124: 1423–1465. <https://doi.org/10.1086/688176>
- Hoberg G, Phillips GM. Text-Based Industry Momentum. *Journal of Financial and Quantitative Analysis*. 2018; 53: 2355–2388. <https://doi.org/10.1017/s0022109018000479>
- Hu D, Zheng H. Does ownership structure affect the degree of corporate financial distress in China? *Journal of Accounting in Emerging Economies*. 2015; 5: 35–50. <https://doi.org/10.1108/jaee-09-2011-0037>
- Hu M, Xiong W, Xu C. Analyst coverage, corporate social responsibility, and firm value: Evidence from China. *Global Finance Journal*. 2021; 50: 100671. <https://doi.org/10.1016/j.gfj.2021.100671>
- Huang P, Jiao Y, Li S. Impact of internal control quality on the information content of social responsibility reports: A study based on text similarity—Evidence from China. *International Journal of Accounting Information Systems*. 2022; 45: 100558. <https://doi.org/10.1016/j.accinf.2022.100558>
- Hussainey K, Walker M. The effects of voluntary disclosure and dividend propensity on prices leading earnings. *Accounting and Business Research*. 2009; 39: 37–55. <https://doi.org/10.1080/00014788.2009.9663348>
- Husted BW, Jamali D, Saffar W. Near and dear? The role of location in CSR engagement. *Strategic Management Journal*. 2016; 37: 2050–2070. <https://doi.org/10.1002/smj.2437>
- Im J, Kim H, Miao L. CEO letters: Hospitality corporate narratives during the COVID-19 pandemic. *International Journal of Hospitality Management*. 2021; 92: 102701. <https://doi.org/10.1016/j.ijhm.2020.102701>
- Jensen MC, Meckling WH. THE NATURE OF MAN*. *Journal of Applied Corporate Finance*. 1994; 7: 4–19. <https://doi.org/10.1111/j.1745-6622.1994.tb00401.x>
- Kao EH, Yeh CC, Wang LH, Fung HG. The relationship between CSR and performance: Evidence in China. *Pacific-Basin Finance Journal*. 2018; 51: 155–170. <https://doi.org/10.1016/j.pacfin.2018.04.006>
- Khan HZ, Bose S, Mollik AT, Harun H. “Green washing” or “authentic effort”? An empirical investigation of the quality of sustainability reporting by banks. *Accounting, Auditing & Accountability Journal*. 2021; 34: 338–369. <https://doi.org/10.1108/aaaj-01-2018-3330>
- Kong D, Shi L, Zhang F. Explain or conceal? Causal language intensity in annual report and stock price crash risk. *Economic Modelling*. 2021; 94: 715–725. <https://doi.org/10.1016/j.econmod.2020.02.013>
- Lang MH, Lundholm RJ. Corporate Disclosure Policy and Analyst Behavior. *The Accounting Review*. 1996; 71: 467–492. <https://doi.org/10.2308/tar-9611271989>
- LARCKER DF, ZAKOLYUKINA AA. Detecting Deceptive Discussions in Conference Calls. *Journal of Accounting Research*. 2012; 50: 495–540. <https://doi.org/10.1111/j.1475-679x.2012.00450.x>
- Leung S, Parker L, Curtis J. Impression management through minimal narrative disclosure in annual reports. *The British Accounting Review*. 2015; 47: 275–289. <https://doi.org/10.1016/j.bar.2015.04.002>
- Li Z, Zou Y, Chen Q. CEOs’ secretarial experience and MD&A disclosure quality: a study based on textual analysis. *Asia-Pacific Journal of Accounting & Economics*. 2023; 30: 1305–1326. <https://doi.org/10.1080/16081625.2021.2003210>
- Lin PT, Li P, Akbar A. Examining the influence of institutional investors on the readability of environmental disclosure in CSR reports of Chinese listed firms. *Corporate Social Responsibility and Environmental Management*. 2024; 31: 1254–1267. <https://doi.org/10.1002/csr.2629>
- Liu Q, Wang J, Chi W. The spillover effects of innovation content disclosure in MD&A. *Pacific-Basin Finance Journal*. 2022; 76: 101879. <https://doi.org/10.1016/j.pacfin.2022.101879>
- LOUGHRAN T, MCDONALD B. Textual Analysis in Accounting and Finance: A Survey. *Journal of Accounting Research*. 2016; 54: 1187–1230. <https://doi.org/10.1111/1475-679x.12123>
- Lu Y, Cahan S, Ma D. Is CSR performance related to disclosure tone in earnings announcements? *Accounting Research Journal*. 2019; 32: 129–147. <https://doi.org/10.1108/arj-05-2016-0059>
- Luo Y. Industrial dynamics and managerial networking in an emerging market: the case of China. *Strategic Management Journal*. 2003; 24: 1315–1327. <https://doi.org/10.1002/smj.363>
- Martínez JI, Stöhr BS, Quiroga BF. Family ownership and firm performance: Evidence from public companies in Chile. *Family Business Review*. 2007; 20: 83–94. <https://doi.org/10.1111/j.1741-6248.2007.00087.x>
- Merkel-Davies DM, Brennan NM. Discretionary Disclosure Strategies in Corporate Narratives: Incremental Information or Impression Management? *Journal of Accounting Literature*. 2007; 26: 116–196.
- Muslu V, Radhakrishnan S, Subramanyam KR, Lim D. Forward-Looking MD&A Disclosures and the Information Environment. *Management Science*. 2015; 61: 931–948. <https://doi.org/10.1287/mnsc.2014.1921>
- Nazari JA, Hrazdil K, Mahmoudian F. Assessing social and environmental performance through narrative complexity in CSR reports. *Journal of Contemporary Accounting & Economics*. 2017; 13: 166–178. <https://doi.org/10.1016/j.jcae.2017.05.002>
- Peterson K, Schmardebeck R, Wilks TJ. The Earnings Quality and Information Processing Effects of Accounting Consistency. *The Accounting Review*. 2015; 90: 2483–2514. <https://doi.org/10.2308/accr-51048>
- Pope S, Peilleux J, El Ouadghiri I, Gomes M. Floodlight or Spotlight? Public Attention and the Selective Disclosure of Environmental Information. *Journal of Management Studies*. 2023; 61: 1230–1265. <https://doi.org/10.1111/joms.12920>

- Porter ME, Kramer MR. Strategy and society: the link between competitive advantage and corporate social responsibility. *Harvard Business Review*. 2006; 84: 78–92, 163.
- Pradhan N, Gyanchandani M, Wadhvani R. A Review on Text Similarity Technique used in IR and its Application. *International Journal of Computer Applications*. 2015; 120: 29–34. <https://doi.org/10.5120/21257-4109>
- Qian A, Zhu D. Financial report similarity and the likelihood of administrative punishment: based on the empirical evidence of textual analysis. *China Journal of Accounting Studies*. 2019; 7: 147–169. <https://doi.org/10.1080/21697213.2019.1642604>
- Rezaee Z, Pouryousof A, Soleimani A. An Examination of Management Tone in the MD&A in the Context of Fraud Pentagon. *Journal of Forensic Accounting Research*. 2024; 9: 167–193. <https://doi.org/10.2308/jfar-2023-039>
- Rim H, Ferguson MAT. Proactive Versus Reactive CSR in a Crisis: An Impression Management Perspective. *International Journal of Business Communication*. 2020; 57: 545–568. <https://doi.org/10.1177/2329488417719835>
- Rugman AM, Verbeke A. Corporate strategies and environmental regulations: an organizing framework. *Strategic Management Journal*. 1998; 19: 363–375. [https://doi.org/10.1002/\(sici\)1097-0266\(199804\)19:43.0.co;2-h](https://doi.org/10.1002/(sici)1097-0266(199804)19:43.0.co;2-h)
- Securities and Exchange Commission (SEC). Financial Reporting Manual—Topic 9: Management’s Discussion and Analysis of Financial Position and Results of Operations (MD&A). 2020. Available at: <https://www.sec.gov/about/divisions-offices/division-corporation-finance/financial-reporting-manual/firm-topic-9> (Accessed: 4 December 2025).
- Shen L, Zhou Y. Can Management Discussion and Analysis Predict Corporate Defaults? An Empirical Analysis Based on the Chinese Stock Market. *Journal of Systems & Management*. 2024; 33: 441–459. <https://xtglxb.sjtu.edu.cn/EN/10.3969/j.issn.1005-2542.2024.02.012>
- Sonpar K, Golden-Biddle K. Using Content Analysis to Elaborate Adolescent Theories of Organization. *Organizational Research Methods*. 2008; 11: 795–814. <https://doi.org/10.1177/1094428106297804>
- Sun Y. Do MD&A Disclosures Help Users Interpret Disproportionate Inventory Increases? *The Accounting Review*. 2010; 85: 1411–1440. <https://doi.org/10.2308/accr.2010.85.4.1411>
- Tam K, Xu Q, Fernando G, Schneible RA. “Tone at the top”: management’s discussion and analysis and audit quality. *Managerial Auditing Journal*. 2023; 38: 602–633. <https://doi.org/10.1108/maj-03-2021-3080>
- Tata J, Prasad S. CSR Communication: An Impression Management Perspective. *Journal of Business Ethics*. 2015; 132: 765–778. <https://doi.org/10.1007/s10551-014-2328-z>
- Timbate L, Park CK. CSR Performance, Financial Reporting, and Investors’ Perception on Financial Reporting. *Sustainability*. 2018; 10: 522. <https://doi.org/10.3390/su10020522>
- Velte P. Do CEO incentives and characteristics influence corporate social responsibility (CSR) and vice versa? A literature review. *Social Responsibility Journal*. 2020; 16: 1293–1323. <https://doi.org/10.1108/srj-04-2019-0145>
- Wang Y, Ye Q, Wang JJ, Wang Y. Earnings manipulation and similarity of annual report disclosure: Evidence from China. *Accounting & Finance*. 2023; 63: 1137–1156. <https://doi.org/10.1111/acfi.13076>
- Wei F, Yan S, Su D, Zhang M. Management tone and corporate social responsibility: Evidence from Chinese listed companies’ involvement in rural revitalization. *Finance Research Letters*. 2024; 65: 105461. <https://doi.org/10.1016/j.frl.2024.105461>
- Wu S, Lin P, Liu Y. Information transparency, analyst forecast and stock price synchronicity. *Journal of Central University of Finance & Economics*. 2013; 9: 63–69.
- Xu D, Huang J, Ren X, Ye M. ESG report textual similarity and stock price synchronicity: Evidence from China. *Pacific-Basin Finance Journal*. 2024; 85: 102343. <https://doi.org/10.1016/j.pacfin.2024.102343>
- Xue LD, Wu GZ, Li ZZ. Incremental Information Disclosure of Key Audit Matters and Management Reports: Research Based on MD&A Text Similarity. *Scientific Decision Making*. 2023; 9: 37–51.
- Yang JH, Liu S. Accounting narratives and impression management on social media. *Accounting and Business Research*. 2017; 47: 673–694. <https://doi.org/10.1080/00014788.2017.1322936>
- Yang M, Bento P, Akbar A. Does CSR Influence Firm Performance Indicators? Evidence from Chinese Pharmaceutical Enterprises. *Sustainability*. 2019; 11: 5656. <https://doi.org/10.3390/su11205656>
- Yu D, Bondi M. A Genre-Based Analysis of Forward-Looking Statements in Corporate Social Responsibility Reports. *Written Communication*. 2019; 36: 379–409. <https://doi.org/10.1177/0741088319841612>
- Zhang R, Chu T, Noronha C, Guan J. Corporate value creation, stock price synchronicity and firm value in China: implications for beyond. *Journal of Accounting in Emerging Economies*. 2022; 12: 566–588. <https://doi.org/10.1108/jaee-06-2021-0212>
- Zhang Z, Liu X, Niu H. Financial crisis early warning of Chinese listed companies based on MD&A text-linguistic feature indicators. *PloS One*. 2023; 18: e0291818. <https://doi.org/10.1371/journal.pone.0291818>
- Zhong M, Xu R, Liao X, Zhang S. Do CSR Ratings Converge in China? A Comparison Between RKS and Hexun Scores. *Sustainability*. 2019; 11: 3921. <https://doi.org/10.3390/su11143921>
- Zhu Y, Zhang R. How local tournament incentives influence ESG disclosure: CEO’s genuinely altruistic or self-serving? *Social Responsibility Journal*. 2024; 20: 383–404. <https://doi.org/10.1108/srj-03-2023-0180>