Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA DOI: 10.46254/AP07.20260241
Published: March 25, 2026

Classifying Impression Management in MD&A Using Text Mining: Evidence from Japanese Listed Firms

Takaaki Kawabe
Course of Engineering
Graduate School of Engineering, Kanagawa University
Kanagawa, Japan
r202470192cb@jindai.jp

Ayuko Komura
Faculty of Business Administration
Toyo University
Tokyo, Japan
komura@toyo.jp

Hirohisa Hirai
Department of Industrial and Management Systems Engineering
Waseda University
Tokyo, Japan
hirai@waseda.jp

Abstract

This study proposes a text-based approach to classifying firms that engage in impression management (IM) in corporate narrative disclosures. Listed firms are required to communicate their operating environment and performance to stakeholders; however, prior accounting research documents that managers may suppress unfavorable information or strategically adjust linguistic features (e.g., tone and readability) to present the firm in a favorable light. As large-scale empirical analyses have proliferated, firms engaging in IM—heterogeneous in form and intensity—are pooled with non-IM firms, limiting the ability of linear models that relate performance to textual attributes to detect IM. We address this limitation by defining several performance indicators and applying text mining to firms' Management's Discussion and Analysis (MD&A) to identify whether each indicator is explicitly mentioned. We then classify firms as engaging in IM based on the consistency between underlying performance and disclosure. For instance, poor performance that is not discussed in the MD&A is indicative of IM through omission. In addition, by examining references to alternative indicators, we detect cases where managers acknowledge weak results but shift emphasis toward more favorable outcomes, indicative of strategic emphasis. The proposed labeling scheme supports more nuanced empirical tests of IM in narrative disclosures.

Keywords
Impression Management, Text Mining, MD&A, Narrative Disclosure, Japanese Listed Firms

1. Introduction
Firms disclose information to support the decision-making of shareholders, investors, and other stakeholders. For example, listed companies are required at the end of each reporting period to disclose not only their financial

performance for that period but also business risks and future prospects in their financial reports, taking into account the managerial environment in which they operate. This disclosure system is intended to reduce information asymmetry between managers and external stakeholders, thereby enabling capital market participants to make more informed decisions. Nevertheless, as agency theory has long suggested, managers may present disclosed information in ways that serve their own interests.

Merkl-Davies and Brennan (2007) classified managers' discretionary disclosure behaviors examined in prior studies into two broad and opposing types: those aimed at reducing information asymmetry between managers and stakeholders and those aimed at exploiting such asymmetry. On this basis, they developed a framework for the latter type, namely impression management (IM). In this framework, managerial IM behavior is divided into two overarching dimensions, concealment and attribution, and further into seven categories of IM strategies. For example, concealment refers to the discretionary manipulation of textual characteristics in disclosure narratives—such as readability, positive or negative wording, and persuasiveness—or of numerical disclosures, including the selection of disclosed information and comparison benchmarks, in order to obscure bad news and highlight good news. Among these textual characteristics, readability and tone in particular have been the focus of numerous empirical studies, especially with the development of information technology and text-mining techniques that have facilitated the analysis of narrative disclosures (Gosselin et al., 2021; Luo & Zhou, 2020).

However, IM research has faced different methodological challenges. For example, as summarized by Merkl-Davies and Brennan (2007), early studies relied largely on multiple regression analyses in which textual characteristics were used as explanatory variables. The measurement of those characteristics depended on human coders, primarily experts, including the authors themselves, and MBA students, partly because text-mining techniques were not yet sufficiently developed. This approach may produce measures that are more conceptually consistent with the intended constructs because coders evaluate documents on the basis of their accumulated knowledge and experience. At the same time, however, it has notable drawbacks: coders must read documents individually, which limits sample size, and coder subjectivity cannot be fully eliminated.

By contrast, recent studies have begun to address these limitations by using text-mining techniques to increase sample size and by adopting more objective and quantitative measures of textual characteristics. Nevertheless, the managerial IM behavior that these studies aim to detect may not always be captured as originally intended. Because these studies rely on large samples, the expected results—namely, evidence of managerial IM behavior—may not emerge depending on the sample selection procedure and the empirical method employed. Specifically, if the proportion of firms adopting IM strategies is substantially smaller than that of firms not adopting such strategies, relationships associated with the minority group may be obscured and remain undetected in the empirical findings. In this sense, although prior studies have improved the efficiency and objectivity of measuring textual characteristics, they have generally tested their hypotheses without explicitly identifying firms that actually adopt IM strategies. As a result, the characteristics of firms belonging to the minority group—namely, those that engage in IM strategies—may not have been adequately captured in analyses based on the full sample.

In light of these arguments, this study aims to propose a method for distinguishing firms that engage in IM behavior from those that do not by using text-mining techniques. Specifically, this study applies the proposed classification to the MD&A sections of Japanese annual securities reports, which correspond to Form 10-K filings in the United States. The reason for focusing on the MD&A section is that it contains managers' explanations of current-period performance, business conditions, and future prospects, and therefore provides a setting in which managerial disclosure strategies are likely to be reflected because of the high degree of discretion in narrative reporting. Focusing on Japanese firms is also meaningful for two reasons. First, empirical research on impression management in narrative disclosure remains limited in the Japanese accounting literature. Second, in recent years, Japanese firms have faced growing pressure to improve ROE and capital efficiency, which makes it more likely that managerial discretion over which performance indicators to emphasize, and which relatively unfavorable indicators to downplay, will be reflected in disclosure narratives.

The proposed method may make it possible to identify firms that adopt IM strategies by shifting the focus of managerial discussion. This, in turn, may allow empirical research on managerial IM strategies in narrative disclosure to be conducted in a more refined manner and in closer alignment with the research objective.

2. Shifting the Focus of Disclosure as Impression Management

Managers may engage in concealment as a form of impression management by shifting the focus of disclosure from unfavorable information to favorable information. According to the framework proposed by Merkl-Davies and Brennan (2007), such behavior falls within the category of concealment and may include, for example, (1) concealing unfavorable information, (2) concealing unfavorable information while simultaneously redirecting attention to favorable information, or (3) disclosing unfavorable information while increasing the amount of favorable information in order to reduce the extent to which the unfavorable information is incorporated into investment decisions. Such concealment does not necessarily require managers to withhold all unfavorable information entirely; rather, it may operate through redirecting stakeholders' attention away from unfavorable performance information and toward more favorable performance indicators. These forms of concealment may also be related to the intentional manipulation of tone and readability in disclosure narratives.

Moreover, managers are likely to adopt different strategies in manipulating the tone and readability of disclosure information depending on whether they shift the focus toward particular performance measures. As a starting point, prior research has shown that disclosure tone is positively associated with stock prices (Henry, 2008; Huang et al., 2014), suggesting that managers have an incentive to make disclosure narratives as positive as possible. In addition, from the perspective of readability, poor firm performance is likely to give managers an incentive to impede the transmission of unfavorable information to the market, thereby reducing readability (i.e., the obfuscation hypothesis; Courtis, 1998; Li, 2008).

In light of the foregoing, when managers shift the focus of disclosure, they may be able to maintain a positive tone in the narrative by avoiding reference to poorly performing metrics and instead directing attention toward more favorable performance indicators. In such cases, if the tone of the disclosure is sufficient to present the information to stakeholders as good news, managers may have little incentive to manipulate readability and may therefore refrain from doing so. By contrast, when managers do not shift the focus of disclosure, unfavorable performance on a given metric must be addressed directly. Under these circumstances, it is relatively difficult for managers to present the disclosure in a positive manner, and the narrative is more likely to become neutral or negative in tone. In that case, for the reasons discussed above, the likelihood of reduced readability is expected to increase.

In prior empirical studies, the relationship between performance and textual characteristics, such as tone and readability, has been examined without taking into account the possibility of IM behavior through the avoidance of unfavorable performance information and the redirection of attention toward other performance measures. This form of focus shifting may be empirically identifiable when a poorly performing metric, such as ROE, receives little or no attention in MD&A, while other performance indicators are discussed instead. Against this background, it may be possible to identify forms of IM behavior that have not been documented in earlier research by developing hypotheses and analyzing this relationship in light of the foregoing discussion.

3. Research Design

This section focuses on the MD&A section of annual securities reports, which is particularly likely to reflect managerial disclosure strategies because of the high degree of discretion involved in its narrative content. Specifically, this section explains an approach for classifying disclosures into six groups based on performance level, whether that performance is explicitly mentioned, and whether other strong performance indicators are discussed at the same time. It then examines whether firms in the classified groups that are presumed to adopt managerial IM strategies actually exhibit the expected textual characteristics, using t-tests and regression analyses based on tone, sentiment, and readability.

3.1 Classification Method

First, a list is compiled of the performance measures and indicators to which managers are likely to pay particular attention when disclosing financial results. It is necessary to identify the measures and indicators that are mentioned in the documents, or in the specific section of the documents, to be analyzed when empirically examining whether managers engage in IM behavior and that managers are likely to regard as especially important in terms of favorable or unfavorable performance outcomes. In this study, therefore, particular attention is given to performance indicators in the MD&A section that managers are likely to consider especially important from the perspective of IM strategies. For example, if managers seek to emphasize the effectiveness and efficiency of their business strategy, they may focus on ROIC, whereas if they seek to emphasize the rate of return delivered to shareholders and investors, they may refer

to ROE. Based on the view that managers are likely to pay particular attention in the MD&A section to managerial efficiency and the rate of earnings attributable to shareholders, this study selects ROE (return on equity) together with four ROIC-related indicators: ROA (return on assets based on operating profit), return on ordinary profit to total assets, return on net income to total assets, and ordinary profit margin.

Second, based on the list compiled in Step 1, the frequency with which these performance indicators appear in the target documents is measured. This makes it possible to determine the extent to which each performance indicator is referred to in the narrative. Third, for each document, disclosures relating to the performance indicators identified in Step 2 are compared with the firm's actual performance in order to assess whether the narrative emphasis is consistent with the underlying results. Through this procedure, each firm can, for example, be assigned to one of four groups with respect to ROE, depending on whether ROE is mentioned and whether ROE performance is strong or poor. Furthermore, when ROE is not mentioned, managers may instead adopt a disclosure strategy that shifts attention toward other favorable performance indicators. To account for this possibility, the classification is extended to six groups by additionally considering whether other performance indicators are mentioned.

Here, the criterion for what constitutes "strong" performance must be defined separately for each performance indicator adopted in the analysis. In other words, it is necessary to consider the threshold below which managers have an incentive to engage in IM behavior when disclosing a given performance indicator. For example, with respect to ROE, if current ROE is lower than that of the previous fiscal year, managers may be perceived by shareholders as lacking managerial ability, which could create an incentive to engage in some form of impression management. Similarly, when a firm reports a loss, managers may disclose this outcome directly in order to fulfill their accountability to shareholders while at the same time altering the manner of expression in the disclosure narrative so as to mitigate any adverse impact on stock prices.

This classification makes it possible to examine managerial IM strategies separately across groups. IM behavior itself is not directly observed; rather, the classification identifies disclosure patterns that are consistent with IM. Broadly speaking, Groups A, B, and C consist of firms with strong ROE performance, whereas Groups D, E, and F consist of firms with poor ROE performance.

Among the strong-performance groups, Group A refers explicitly to ROE and may therefore reflect straightforward emphasis on favorable performance. Groups B and C do not refer to ROE and may instead reflect either a routine tendency not to discuss ROE or attention to performance indicators other than those selected in this study.

The distinction between Groups D, E, and F is central to the analysis. Group D consists of firms with poor ROE performance that nonetheless refer to ROE. These disclosures may reflect either candid reporting or impression management through the way unfavorable performance is linguistically framed. Group E consists of firms with poor ROE performance that do not refer to ROE but do refer to other performance indicators. This group is most consistent with impression management through shifting the focus of disclosure away from unfavorable ROE and toward relatively more favorable indicators. Group F consists of firms with poor ROE performance that refer neither to ROE nor to other performance indicators. These firms may simply lack a routine practice of disclosing ROE, or they may adopt a strategy of silence by withholding unfavorable information from stakeholders (Leung et al., 2015).

Here, the criterion for determining whether a performance indicator is "mentioned" is whether the indicator name itself appears explicitly in the text or whether an equivalent expression is used in paraphrased form. Accordingly, this study does not take into account the extent to which the indicator is discussed in depth or analytically interpreted. This is also a limitation of the bag-of-words approach adopted in this study. In particular, with respect to paraphrased expressions, references to ROE are identified using three patterns—"ROE," "return on equity," and "return on net income to shareholders' equity"—whereas references to ROA are identified using three patterns—"ROA," "return on assets based on operating profit," and "return on assets."

3.2 Sample Selection

The sample selection procedure is conducted in two stages. First, this study examines annual securities reports submitted through EDINET by firms whose fiscal year-end fell between January 1, 2020 and December 31, 2024, resulting in an initial sample of 21,217 firm-year observations. After matching these documents with financial data from Nikkei NEEDS FinancialQUEST and excluding non-listed firms and observations with missing performance-

indicator data, the classification procedure is applied to 17,874 firm-year observations. The empirical tests focus on Groups D and E, yielding 1,454 Group D and 895 Group E documents under the prior-period ROE benchmark and 260 Group D and 214 Group E documents under the 0% benchmark. After excluding observations with missing regression variables, the final regression samples consist of 2,349 and 470 firm-year observations, respectively.

3.3 Hypothesis Development and Testing Method

In empirically examining managerial IM strategies, a more refined analysis may be possible by first applying the classification developed in Section 3.1 and then testing whether firms within a more narrowly defined sample actually adopt such strategies. In this study, the analysis focuses on Groups D and E under this framework in order to examine whether managers in these groups in fact engage in the IM strategy discussed in Section 2, namely, the redirection of attention away from the focal performance measure. The reason for focusing on Groups D and E is as follows. Among Groups D, E, and F, all of which exhibit poor performance and therefore are likely to face managerial incentives for IM behavior, firms classified into Group F may follow a strategy different from the redirection-based IM strategy described in Section 2, or may simply lack a routine practice of referring to the specific indicators selected in this study. For this reason, Group F is excluded from the analysis. Thus, the empirical analysis compares Groups D and E to assess whether attention redirection is associated with distinct textual characteristics.

From the perspective of IM strategies based on redirecting attention away from the focal performance measure, Group D may be interpreted as consisting of documents in which such redirection does not occur, because ROE is explicitly mentioned despite poor ROE performance. Nevertheless, as discussed in Section 2, managers may instead adopt an IM strategy through the manipulation of textual characteristics. More specifically, because these disclosures refer to poor ROE performance—that is, they contain unfavorable information—their tone or sentiment is expected to be more negative or, at minimum, closer to neutral. Accordingly, managers may instead have an incentive to adopt an IM strategy that lowers readability.

By contrast, Group E consists of documents in which ROE is not mentioned despite poor ROE performance, whereas other performance indicators are discussed instead. In these documents, managers may refrain from referring to the unfavorable ROE and instead selectively highlight other performance indicators that are relatively more favorable. As a result, disclosures in Group E may exhibit a more positive tone than those in Group D. With respect to readability, if managers judge that the positive nature of the disclosure is already sufficient to appeal to stakeholders, the incentive to reduce readability may be weaker. The readability of documents in Group E may therefore be higher than that of documents in Group D. On this basis, the following hypotheses are proposed.

H1: Documents classified into Group E exhibit a more positive tone than documents classified into Group D.
H2: Documents classified into Group E exhibit higher readability than documents classified into Group D.

3.4 Empirical Model

To test the hypotheses, the following empirical model is estimated, where t denotes the fiscal year of the annual securities report:

$$\begin{aligned} Text\ Character_t = {} & \beta_0 + \beta_1 \cdot Edummy_t + \beta_2 \cdot SIZE_t + \beta_3 \cdot EARNVOL_t + \beta_4 \cdot ROAVOL_t \\ & + \beta_5 \cdot d2021 + \beta_6 \cdot d2022 + \beta_7 \cdot d2023 + \beta_8 \cdot d2024 + \varepsilon \end{aligned} \quad (1)$$

Here, Text Character includes the two tone measures, the two sentiment measures, and the readability-related variable described in Section 3.5. Edummy indicates that a document belongs to Group E. As control variables, SIZE is measured as the natural logarithm of market capitalization, EARNVOL is the standard deviation of return on net income to total assets over the previous five years, ROAVOL is the standard deviation of ROA over the previous five years, and d2021 through d2024 are year dummy variables that take the value of one when the observation belongs to the corresponding fiscal year. These variables are selected from those used in Henry and Leone (2016) and Kato and Goto (2021), subject to data availability and the need to secure a sufficient sample size. In addition, all explanatory variables except dummy variables and Tone1 are winsorized at the upper and lower 2.5 percent levels and then standardized.

3.5 Variable Construction

This study adopts a bag-of-words approach for text mining. When applying this approach to Japanese text, unlike English, words are not separated by spaces, and it is therefore necessary to segment sentences into the smallest lexical units, namely morphemes. For this purpose, the morphological analyzer MeCab is employed. However, the default system dictionary included in MeCab (the IPA dictionary, which specifies word segmentation and corresponding priority rules) has difficulty identifying compound expressions specific to accounting documents. To address this issue, compound terms consistent with the accounting context were extracted using the Python library termextract from annual securities reports submitted over the five-year period from July 1, 2016 to June 30, 2021, and these terms were incorporated into the system dictionary so that they would be processed with priority. Through this procedure, the text segmentation is made more consistent with the accounting context.

As textual characteristics, this study employs tone—that is, indicators of whether the text is tilted toward a positive or negative orientation—and sentiment, which captures the polarity of the text. Specifically, two tone measures, Tone1 and Tone2, are defined by Equations (2) and (3), and two sentiment measures, PosSent and NegSent, are defined by Equations (4) and (5). Tone indicates whether a given narrative is relatively more positive or more negative overall. By contrast, sentiment captures the strength of polarity in the text, where PosSent (NegSent) indicates the extent to which the text exhibits positive (negative) characteristics. By distinguishing between tone and sentiment, the analysis may provide more fine-grained insights into whether managers adjust disclosure while balancing positive and negative characteristics in the text, or whether they manage disclosure by selectively emphasizing one polarity over the other (Kawabe et al., 2026).

$$Tone_1 = \frac{Pos - Neg}{Pos + Neg} \tag{2}$$

$$Tone_2 = \frac{Pos - Neg}{Total\ Words} \tag{3}$$

$$PosSent = \frac{Pos}{Total\ Words} \tag{4}$$

$$NegSent = \frac{Neg}{Total\ Words} \tag{5}$$

Here, Pos (Neg) denotes the frequency of positive (negative) words, and TotalWords denotes the total number of words in the text. Positive and negative words are identified using the accounting-specific sentiment dictionary proposed by Hirai et al. (2024). This dictionary is a Japanese word list designed to classify the polarity of disclosure-related language, developed with reference to the method used by Loughran and McDonald (2011) to construct a polarity dictionary for corporate disclosure documents. Because it was developed on the basis of annual securities reports, it is better suited to the context of accounting documents than general-purpose sentiment dictionaries developed in psychology or linguistics (Hirai et al., 2024).

As a readability-related measure, this study uses Length, defined as the logarithm of the total number of words. Li (2008) employed document length as an indicator capturing processing burden arising from narrative extensiveness in annual reports, and this measure has also been used in prior studies on Japanese disclosure documents. However, Length should be interpreted not as a comprehensive measure of readability itself, but rather as a parsimonious proxy for the amount of text and the degree of verbosity. Accordingly, in this study, a larger value of Length is interpreted as indicating a higher information-processing cost for readers, at least from the perspective of document length.

$$Length = Log(Total\ Words) \tag{6}$$

That said, it is difficult to argue that a unified standard measure of readability has been established for accounting documents. Loughran and McDonald (2014) pointed out that the widely used Fog index does not function adequately in financial texts because multisyllabic words do not necessarily correspond to difficult words in such contexts. In addition, quantity-based measures such as word count and file size capture only one aspect of readability, namely document length, and do not provide a comprehensive representation of readability. More recently, measures such as the Bog Index, which incorporates plain-English attributes, and multidimensional measures of clarity and conciseness have been proposed (Bonsall et al., 2017; Efretuei and Yekini, 2025). Readability research has therefore evolved from simple measures based on document length or Fog-type indices toward more context-dependent and multidimensional

approaches. Against this background, the value of using Length in this study lies in its ability to capture, within the broader concept of readability in accounting documents, the specific aspects of verbosity and information-processing cost. As in the international literature, the development of readability measures better suited to Japanese accounting documents remains an urgent task.

4. Results and discussion

4.1 Classification Results

First, the results of the difference tests in average mention shares for each performance indicator for each performance indicator by favorable versus unfavorable performance are reported in Table 1. The values shown for each indicator represent the average mention share for that indicator. The symbols ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 1. Results of Welch's t-Tests for Differences in Average Mention Shares by Performance Classification

Panel A: Results of Welch's t-test when prior-period performance is used as the benchmark.

	n	ROE		ROA		OrdProfit / Assets		NetProfit /Assets		OrdProfit /Sales	
A, B, C	8,412	0.364		0.034		0.024		0.002		0.102	
D, E, F	8,814	0.337		0.037		0.027		0.002		0.102	
Differences		0.027	**	-0.003		-0.003		0.000		0.001	

Panel B: Results of Welch's t-test when whether ROE exceeds 0% is used as the benchmark.

	n	ROE		ROA		OrdProfit / Assets		NetProfit /Assets		OrdProfit /Sales	
A, B, C	15,080	0.370		0.034		0.027		0.003		0.107	
D, E, F	2,747	0.192		0.036		0.016		0.000		0.068	
Differences		0.178	***	-0.002		0.011	***	0.003	***	0.038	

When prior-period change is used as the benchmark for the performance criterion managers are likely to consider, a statistically significant difference is observed only for ROE at the 5% level, with mention shares being higher for strong-performance firms than for poor-performance firms. This suggests that managers of firms with relatively poor ROE may refrain from referring to ROE compared with managers of firms whose ROE is more favorable. When whether ROE exceeds 0% is used as the benchmark, statistically significant differences at the 1% level are observed for ROE, return on ordinary profit to total assets, return on net income to total assets, and ordinary profit margin, again with mention shares being higher for favorable-performance firms than for unfavorable-performance firms. These results suggest that poor performance is negatively associated with references to the corresponding performance indicators in the MD&A section. Taken together, the findings are consistent with the possibility that managers selectively reduce references to performance indicators when firm performance is unfavorable.

Next, based on the classification procedure described in Section 3.1, the documents selected in Section 3.2 are classified into six groups (Figure 1). ROE is used as the focal performance metric for identifying impression management strategies, and two criteria are considered: whether current ROE exceeds prior-period ROE and whether current ROE exceeds 0%. The documents are then assigned to groups according to whether ROE satisfies or falls below each benchmark, whether ROE is mentioned, and, when ROE is not mentioned, whether other performance indicators are mentioned instead.

Using the procedure in Section 3.1, the sample is classified into six groups under two ROE benchmarks. Group E, which is interpreted as the group most consistent with focus-shifting impression management, accounts for 895 of 17,226 observations under the prior-period benchmark and 214 of 17,874 under the 0% benchmark. These

Panel B: Six-group classification results when whether ROE exceeds 0% is used as the benchmark.

	ROE Mentioned	ROE Not Mentioned	
		Other Performance Mentioned	Other Performance Not Mentioned
$ROE_t > 0\%$	**A: 2,724**	**B: 1,574**	**C: 10,817**
$ROE_t < 0\%$	**D: 260**	**E: 214**	**F: 2,285**

Panel A: Six-group classification results when prior-period ROE is used as the benchmark.

	ROE Mentioned	ROE Not Mentioned	
		Other Performance Mentioned	Other Performance Not Mentioned
$ROE_t > ROE_{t-1}$	**A: 1,486**	**B: 825**	**C: 6,101**
$ROE_t < ROE_{t-1}$	**D: 1,454**	**E: 895**	**F: 6,465**

Figure 1. Six-Group Classification Results under Alternative ROE Benchmarks

figures indicate that such firms constitute only a minority of the full sample, which supports the use of a classification-based approach in the subsequent regression analysis.

4.2 Descriptive Statistics

Descriptive statistics for the classified samples are reported in Table 2. Under both benchmarks, the mean values of Tone1 and Tone2 are above zero, and the mean of PosSent exceeds that of NegSent. These patterns suggest that even in samples consisting of firms classified as having poor performance, MD&A narratives are not, on average, strongly tilted toward negative expression. At the same time, Panel B, which uses the 0% ROE criterion, exhibits lower tone and higher negative sentiment than Panel A, suggesting that the narratives become relatively more negative when the benchmark captures more severe poor performance. Turning to Length, its mean is 8.444 in Panel A and 8.448 in Panel B, indicating no substantial difference in document length across the two samples. Thus, although the Panel B

Table 2. Descriptive Statistics for the Classified Samples

Panel A: Descriptive statistics for the sample when prior-period performance is used as the benchmark.

	N	Mean	Median	SD	Max	Min
$Tone_{1,t}$	**2,349**	**0.277**	**0.289**	**0.274**	**0.939**	**-0.579**
$Tone_{2,t}$	**2,349**	**0.011**	**0.011**	**0.011**	**0.044**	**-0.031**
$PosSent_t$	**2,349**	**0.026**	**0.025**	**0.007**	**0.051**	**0.002**
$NegSent_t$	**2,349**	**0.015**	**0.014**	**0.006**	**0.051**	**0.001**
$Length_t$	**2,349**	**8.444**	**8.429**	**0.337**	**9.900**	**7.398**
$Edummy_t$	**2,349**	**0.381**	**0**	**0.486**	**1**	**0**
ROE_t	**2,349**	**-0.038**	**0.058**	**1.575**	**1.594**	**-65.095**
$SIZE_t$	**2,349**	**10.420**	**10.315**	**0.780**	**13.272**	**8.619**
$EARNVOL_t$	**2,349**	**0.022**	**0.012**	**0.041**	**1.045**	**0.000**
$ROAVOL_t$	**2,349**	**0.015**	**0.008**	**0.031**	**0.907**	**0.000**

Panel B: Descriptive statistics for the sample when whether ROE exceeds 0% is used as the benchmark.

	N	Mean	Median	SD	Max	Min
Tone $_{1,t}$	470	0.104	0.102	0.252	0.727	-0.543
Tone $_{2,t}$	470	0.004	0.004	0.010	0.033	-0.031
PosSent $_t$	470	0.023	0.022	0.006	0.042	0.008
NegSent $_t$	470	0.019	0.018	0.007	0.051	0.003
Length $_t$	470	8.448	8.429	0.311	9.361	7.582
Edummy $_t$	470	0.449	0	0.498	1	0
ROE $_t$	470	-0.591	-0.082	3.506	0.000	-65.095
SIZE $_t$	470	10.037	9.897	0.667	12.455	8.619
EARNVOL $_t$	470	0.049	0.024	0.080	0.715	0.000
ROAVOL $_t$	470	0.028	0.013	0.049	0.416	0.000

Table 3.. Multiple Regression Results for Textual Characteristics under Alternative ROE Benchmarks

Panel A: Multiple regression results when prior-year ROE is used as the benchmark for the IM strategy.

Dep.	Tone $_{1,t}$			Tone $_{2,t}$			PosSent $_t$			NegSent $_t$			Length		
	coef.	t-value	p-value	coef.	t-value	p-value	coef.	t-value	p-value	coef.	t-value	p-value	coef.	t-value	p-value
const	0.1587	13.709	0.000 ***	0.0065	13.494	0.000 ***	0.0239	80.419	0.000 ***	0.0174	60.726	0.000 ***	8.5132	623.444	0.000 ***
Edummy	0.0212	1.972	0.049 **	0.0003	0.778	0.436	-0.0001	-0.494	0.622	-0.0005	-1.847	0.065 *	-0.0611	-4.426	0.000 ***
ROE $_t$	0.0849	13.801	0.000 ***	0.0033	13.576	0.000 ***	0.0014	10.676	0.000 ***	-0.0018	-11.754	0.000 ***	-0.0271	-3.913	0.000 ***
SIZE $_t$	0.0245	4.415	0.000 ***	0.0012	5.026	0.000 ***	0.0008	5.047	0.000 ***	-0.0004	-2.942	0.003 ***	0.0907	11.609	0.000 ***
EARNVOL $_t$	0.0221	2.249	0.025 **	0.0005	1.310	0.190	-0.0003	-1.281	0.200	-0.0008	-3.456	0.001 ***	0.0056	0.458	0.647
ROAVOL $_t$	-0.0099	-1.031	0.303	-0.0003	-0.687	0.492	0.0001	0.519	0.604	0.0004	1.706	0.088 *	0.0078	0.653	0.514
Year Dummy	**Included**			**Included**			**Included**			**Included**			**Included**		
N	2.349			2.349			2.349			2.349			2.349		
Adj. R-squared	0.208			0.195			0.113			0.174			0.095		

Panel B: Multiple regression results when whether ROE exceeds 0% is used as the benchmark for the IM strategy.

Dep.	Tone $_{1,t}$			Tone $_{2,t}$			PosSent $_t$			NegSent $_t$			Length		
	coef.	t-value	p-value	coef.	t-value	p-value	coef.	t-value	p-value	coef.	t-value	p-value	coef.	t-value	p-value
const	0.0016	0.061	0.952	-0.0001	-0.056	0.956	0.0210	34.904	0.000 ***	0.0210	29.465	0.000	8.4770	281.442	0.000 ***
Edummy	-0.0135	-0.569	0.570	-0.0009	-0.941	0.347	-0.0005	-0.764	0.445	0.0005	0.690	0.490	0.0093	0.331	0.741
ROE $_t$	0.0136	1.111	0.266	0.0006	1.293	0.196	0.0008	3.024	0.002 ***	0.0002	0.670	0.503	-0.0519	-3.155	0.002 ***
SIZE $_t$	0.0217	1.813	0.070 *	0.0010	1.916	0.055 *	0.0008	2.519	0.012 *	-0.0002	-0.651	0.515	0.0988	7.036	0.000 ***
EARNVOL $_t$	0.0386	1.732	0.083 *	0.0017	2.027	0.043 **	0.0008	1.613	0.107	-0.0009	-1.582	0.114	-0.0005	-0.022	0.982
ROAVOL $_t$	-0.0100	-0.423	0.672	-0.0005	-0.668	0.504	-0.0006	-1.389	0.165	-0.0001	-0.181	0.857	-0.0323	-1.560	0.119
Year Dummy	**Included**			**Included**			**Included**			**Included**			**Included**		
N	470			470			470			470			470		
Adj. R-squared	0.120			0.126			0.105			0.087			0.111		

narratives appear relatively more negative, they cannot be regarded as more concise, at least from the perspective of document length.

4.3 Regression Results

The regression results are reported in Table 3. When prior-period ROE is used as the benchmark, ROE t is significantly associated with all tone and sentiment measures in the expected directions, indicating that better performance is associated with more positive and less negative narrative expression. When the 0% ROE criterion is used, only PosSent remains significantly related to ROE t. Overall, these findings suggest that MD&A narratives partly reflect underlying performance, although negative wording does not appear to increase proportionately with poor performance.

ROE t is negatively associated with Length under both benchmarks, indicating that poorer performance is associated with longer MD&A narratives. This pattern is consistent with prior evidence that low-performing firms issue longer, potentially harder-to-process annual reports and may reflect either additional explanation or greater verbosity.

When prior-period ROE is used as the benchmark, Edummy is associated with higher Tone1, lower NegSent, and shorter Length, whereas these relations are not significant under the 0% benchmark. These results provide partial support for H1 and H2 and are consistent with the interpretation that Group E firms weaken the negative impression of poor ROE not by lengthening the text, but by shifting the focus of disclosure toward other performance indicators. Next, the interpretation focuses on Edummy. With respect to tone and sentiment, when prior-period ROE is used as the benchmark, Edummy is significantly associated with Tone1 and NegSent (Tone1: coef. = 0.021, $p < 0.05$; NegSent: coef. = -0.0005, $p < 0.10$). This suggests that the documents in Group E exhibit relatively weaker negative polarity than those in Group D, which in turn results in a higher tone. This finding is consistent with the argument developed in Sections 2 and 3.3 that managers may shift attention away from poor performance and toward other indicators, thereby reducing the relative frequency of negative words.

However, when the 0% ROE criterion is used, Edummy is not significantly related to any of the tone or sentiment measures. Hypothesis 1 is therefore interpreted as being partially supported.

Further interpretation can be obtained from the relationship between Edummy and the readability-related measure. When prior-period ROE is used as the benchmark, Edummy is found to have a statistically significant negative coefficient on Length (coef. = -0.0611, $p < 0.01$), whereas no such tendency is observed when the 0% ROE criterion is used. Thus, the documents in Group E tend, in part, to have a smaller amount of text in the MD&A section than those in Group D. It should be noted, however, that the Length measure used in this study captures only the amount of text and does not directly measure readability as a whole. Accordingly, this result suggests only that, from the perspective of Length, documents in Group E may be less verbose than those in Group D. Hypothesis 2 is therefore also regarded as partially supported.

Taken together with the relationship between ROE_t and Length, these results indicate that managers of poorly performing firms tend to produce longer MD&A narratives. This may be interpreted as an effort to gain stakeholder understanding by providing more extensive discussion of the firm's business situation. At the same time, however, such lengthening may also represent more verbose disclosure intended to increase stakeholders' information-processing costs and thereby reduce the adverse effects of poor performance on stock prices or managerial employment risk. By contrast, documents in Group E are shown to be shorter and to contain fewer negative words than those in Group D. Taken together, these findings suggest that impression management in Group E may not operate through lengthening the text and increasing processing costs, but rather through shifting the focus of disclosure away from ROE and thereby weakening the relative negative impression conveyed to stakeholders. In other words, among documents classified into Group E, impression management appears to be associated less with obfuscation through increased textual length and more with the selection of what to mention and the reallocation of emphasis across performance indicators.

5. Conclusion

This study proposed a six-group classification framework based on whether a focal performance indicator is mentioned and whether other performance indicators are discussed, and examined its usefulness for identifying managerial impression management in the MD&A section using Japanese annual securities reports. The framework is designed

to identify, in advance, document groups likely to reflect impression management through shifts in disclosure focus, particularly when such firms constitute only a minority of the full sample.

The empirical results show that lower performance is associated with longer MD&A narratives among poorly performing firms as a whole. By contrast, firms in Group E, which are likely to adopt impression management by shifting the focus of disclosure, exhibit, in some specifications, higher tone, lower negative sentiment, and shorter text than firms in Group D. These findings suggest that impression management in Group E operates less through textual lengthening and more through shifting the focus of disclosure toward performance indicators other than ROE.Taken together, these findings suggest that impression management in corporate disclosure is not necessarily uniform and may take multiple forms. In particular, the evidence from the MD&A section indicates that impression management may occur both through strategies involving textual lengthening and reduced readability, and through strategies involving the selective choice of disclosure targets and changes in emphasis. In this respect, the present study contributes to the literature on tone and readability by introducing the perspective of not only how managers write, but also what they choose to write about and what they do not discuss. The proposed classification approach may therefore serve as a useful tool for more precisely identifying firms that are likely to adopt impression management and for examining managerial behavior in narrative disclosure.

At the same time, this study has several limitations. First, although documents assigned to Group E are interpreted as belonging to a group likely to adopt impression management through shifts in focus, not all such documents necessarily reflect actual impression management. Firms may omit reference to ROE for reasons such as disclosure routines or managerial preferences for performance indicators other than ROE. Second, the performance indicators considered in this study are limited to ROE and several ROIC-related measures, and the classification results may change if additional financial indicators are incorporated. Third, the readability-related measure used in this study, Length, captures only one aspect of readability—namely, document length—and does not fully represent readability in Japanese MD&A narratives. Fourth, this study relies on observational data, and although the observed relationships are consistent with impression management, they do not directly reveal managerial intent or the underlying causal mechanisms.

The MD&A section of annual securities reports is where managers analyze and disclose the firm's performance and business conditions for the relevant fiscal period, and its content and expression involve a substantial degree of managerial discretion.Several directions for future research follow from these limitations. First, the classification framework could be extended by incorporating a broader set of performance indicators. It may also be possible to identify impression-management firms more precisely by incorporating into the classification algorithm whether the referenced performance indicators are themselves favorable and whether firms have a routine practice of referring to the selected indicators. Second, the identification of impression management through shifts in focus could be improved by refining readability measures suitable for Japanese disclosure documents and by validating the classification through annotation-based approaches. Third, it would be desirable to examine how the documents identified in this study affect market participants by linking them to market-based outcomes such as stock-price reactions, analyst forecasts, and the cost of capital. More broadly, it would also be important to examine the generalizability of the present findings by extending the analysis to other institutional settings and to other forms of narrative disclosure, such as earnings presentation materials, integrated reports, and earnings summaries.

References

Bonsall, S. B., Leone, A. J., Miller, B. P., and Rennekamp, K., "A Plain English Measure of Financial Reporting Readability," *Journal of Accounting and Economics*, vol. 63, no. 2–3, pp. 329–357, 2017. https://doi.org/10.1016/j.jacceco.2017.03.002

Courtis, J. K., "Annual Report Readability Variability: Tests of the Obfuscation Hypothesis," *Accounting, Auditing & Accountability Journal*, vol. 11, no. 4, pp. 459–471, 1998.

Efretuei, E., and Yekini, K., "Clarity and Conciseness of Financial Narrative Disclosures: Does the Practice Statement Matter?," *Abacus*, vol. 61, no. 4, pp. 1110–1144, 2025. https://doi.org/10.1111/abac.70018

Gosselin, M., Le Maux, J., and Smaili, N., "Readability of Accounting Disclosures: A Comprehensive Review and Research Agenda," *Journal of Accounting Literature*, vol. 48, pp. 167–197, 2021. https://doi.org/10.1016/j.acclit.2021.07.002

Henry, E., "Are Investors Influenced by How Earnings Press Releases Are Written?," *The Journal of Business Communication*, vol. 45, no. 4, pp. 363–407, 2008. https://doi.org/10.1177/0021943608319388

Henry, E., and Leone, J. A., "Measuring Qualitative Information in Capital Markets Research: Comparison of Alternative Methodologies to Measure Disclosure Tone," *The Accounting Review*, vol. 91, no. 1, pp. 153–178, 2016. https://doi.org/10.2308/accr-51161

Hirohisa, H., Ayuko, K., and Takaaki, K., "Applicability of an Accounting-Specific Polarity Dictionary in Tone Analysis," *Sangyo Keiri*, vol. 83, no. 4, pp. 41–50, 2024.

Huang, X., Teoh, S. H., and Zhang, Y., "Tone Management," *The Accounting Review*, vol. 89, no. 3, pp. 1083–1113, 2014. https://doi.org/10.2308/accr-50684

Daisuke, K., and Keiichi, G., "Text Analysis of Annual Securities Reports: Managerial Forward-Looking Disclosure and Future Firm Performance," *Kin'yu Kenkyu*, vol. 40, no. 3, pp. 45–75, 2021.

Kawabe, T., Komura, A., and Hirai, H., "Comparison of Tone and Sentiment Measurements for Japanese Corporate Disclosures," *Journal of the Japan Industrial Management Association*, vol. 76, no. 4, pp. 138–145, 2026.

Leung, S., Parker, L., and Courtis, J., "Impression Management Through Minimal Narrative Disclosure in Annual Reports," *The British Accounting Review*, vol. 47, no. 3, pp. 275–289, 2015.

Li, F., "Annual Report Readability, Current Earnings, and Earnings Persistence," *Journal of Accounting and Economics*, vol. 45, no. 2–3, pp. 221–247, 2008. https://doi.org/10.1016/j.jacceco.2008.02.003

Loughran, T., and McDonald, B., "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks," *The Journal of Finance*, vol. 66, no. 1, pp. 35–65, 2011. https://doi.org/10.1111/j.1540-6261.2010.01625.x

Loughran, T., and McDonald, B., "Measuring Readability in Financial Disclosures," *The Journal of Finance*, vol. 69, no. 4, pp. 1643–1671, 2014. https://doi.org/10.1111/jofi.12162

Luo, J.-H., and Zhou, D., "Corporate Financial Disclosure Tone: A Survey of the Literature," *Review of Quantitative Finance and Accounting*, vol. 55, no. 3, pp. 1001–1026, 2020. https://doi.org/10.1007/s11156-020-00888-6

Merkl-Davies, D. M., and Brennan, N. M., "Discretionary Disclosure Strategies in Corporate Narratives: Incremental Information or Impression Management?," *Journal of Accounting Literature*, vol. 26, pp. 116–196, 2007.

Biographies

Takaaki Kawabe is a PhD student of the course of engineering at the Graduate School of Engineering at Kanagawa University, Japan.

Ayuko Komura is an associate professor at the Faculty of Business Administration at Toyo University, Japan. Her research topics include operating profit stability and management controls. She earned her PhD in Business Administration at Meiji University. She received a fellowship from the Japan Society for the Promotion of Science between 2017 and 2019.

Hirohisa Hirai is a professor at the Department of Industrial and Management Systems Engineering at Waseda University, Japan. He earned his PhD in Engineering at Osaka University. His research is an empirical study of management accounting, firm analysis, and applied statistics, particularly firm valuations. He has received awards from several academic societies in accounting and management. These forms of concealment may also relate the intentional manipulation These forms of concealment may also relate the intentional manipulation