STRATEGIC NARRATIVES IN REGULATED DISCLOSURES: CONTENT, DYNAMICS, AND COORDINATION

MINQI LIU

A DISSERTATION SUBMITTED TO
THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN ADMINISTRATION
YORK UNIVERSITY
TORONTO, ONTARIO, CANADA

AUGUST 2026

© MINQI LIU, 2026

Abstract

Corporate ESG disclosures are not neutral transmissions of information. They are strategic constructions in which managers decide what to say, how long to sustain it, and where to position it across the documentary architecture of corporate reporting. This dissertation presents three empirical studies that examine these choices in regulated disclosure settings. The first applies structural topic modeling and LASSO regression to 1,980,034 sentences from 13,057 compensation discussion & analysis filings by S&P 1500 firms (2006–2015), finding that changes in accounting performance are systematically associated with differences in topic emphasis. The second develops a sentence-level disclosure lifecycle framework and applies FinBERT-ESG embeddings and BERTopic to 80,890 voluntary social disclosure sentences from 12,774 Canadian management information circulars (1997–2023), finding that dominant social topics follow distinct patterns of being introduced, carried forward, and abandoned over time. The third classifies social and environmental sentences in 3,416 matched annual information forms (AIFs)–management discussion & analysis (MD&A) filing pairs from Canadian mining firms (1997–2024) into overlapping and outlet-specific disclosures, finding that each documentary position is associated with a distinct accountability logic. Collectively, these studies shed light on variation in what firms disclose, how disclosure content evolves over time, and how particular accounts are positioned across reporting outlets, as well as the strategic and institutional contexts associated with these patterns.

Dedication

To my husband, Haoyu, my beloved companion and safe harbor. Thank you for walking beside me through every stage of this journey. And to my daughter, Youyan, who brought immeasurable joy into my life while this dissertation was being written. Completing this work has been all the more meaningful because of both of you.

Acknowledgements

My deepest thanks go to my supervisor and mentor, Dr. Dean Neu, whose trust gave me the freedom to develop my own ideas and whose guidance helped me grow as a researcher. His generosity, scholarly insight, and patience have shaped both this dissertation and my development throughout the doctoral program.

I owe equally sincere thanks to Dr. Gregory Saxton, whose encouragement, methodological rigor, and thoughtful feedback strengthened this dissertation at every stage. Greg's generosity with his time and his willingness to engage seriously with new ideas made the research process more rewarding.

I am also grateful to Dr. Kiridaran Kanagaretnam and Dr. Nelson Waweru for serving on my committee and for their valuable feedback. I would also like to thank my external examiner, Dr. Jamal Nazari, for his thoughtful comments and insightful suggestions. Beyond my committee, I would like to thank the faculty, staff, and fellow doctoral students at the Schulich School of Business for their support, encouragement, and friendship during my years in the PhD program.

I would also like to acknowledge my family. To my husband, Haoyu, my companion through every late night, setback, and small victory, thank you for your patience, humor, and love. To my parents, thank you for your constant support and belief in me. Your encouragement has meant more than words can express. Finally, to our daughter, Youyan: you arrived in the middle of this dissertation and brought a new perspective to it. As these chapters gradually took shape, so did you. Watching you grow alongside this work has been a privilege and a joy.

Table of Contents

Abstract ii
Dedication iii
Acknowledgements iv
Table of Contents v
List of Tables viii
List of Figures ix
Chapter 1: Introduction 1
Chapter 2: Accounting Numbers and Executive Compensation Narratives 8
1. Introduction 9
2. Background 12
3. Theoretical Framework 14
3.1 Topics in CD&A Disclosures 14
3.2 Changes in Accounting Performance and Choice of CD&A Topics 15
4. Data and Method 16
4.1 Sample Selection 16
4.2 Measuring CD&A Topics 18
4.3 Measure of Changes in Accounting Performance Numbers 21
4.4 Control Variables 22
4.5 LASSO Regression Model 22
5. Findings 24
5.1 CD&A Topics 24
5.2 CD&A topics and Changes in Accounting Performance 30
6. Additional Analyses and Robustness Tests 37
6.1 Miscellaneous Robustness Tests 37
6.2 Alternative Measure of Topic 37
6.3 Results of Filing-Level LASSO Regressions 38
6.4 Alternative Measures of Accounting Number Change 43
6.5 Relationship between CEO Pay Raises and Topic Choice 43
6.6 Earnings Management and CD&A Topic Adoption 45

6.7 CD&A Topic Choices and Market Reactions 47
6.8 Topics and Future Accounting Number Change 49
7. Discussion 50
Chapter 3: The Dynamics of Voluntary Social Disclosures in Management Information Circulars 53
1. Introduction 54
2. Theoretical Framework 59
2.1 Visibility 59
2.2 Temporality 60
2.3 Institutional Expectations 61
3. Data and Method 63
3.1 Sample Selection 63
3.2 Topic Modeling 66
3.3 Lifecycle Classification 67
3.4 Analysis of Lifecycle Patterns 68
3.5 Institutional Expectations and Lifecycle Outcomes 70
4. Findings 71
4.1 Voluntary Social Topics 71
4.2 Topic Content and Disclosure Lifecycle Dynamics 73
4.3 Institutional Expectations and Lifecycle Outcomes 79
5. Discussion 82
Chapter 4: Recurring and Outlet-Specific Social and Environmental Accounts in Corporate Reporting 87
1. Introduction 88
2. Theoretical Framing 93
2.1 Corporate Reporting as a Calculative Practice 93
2.2 Inter-Documentary Ordering of S&E Disclosure 94
2.3 Accountability Relations and Organizing Logics of S&E Disclosure 96
3. Data and Method 98
3.1 Sample 98
3.2 Identifying Overlapping and Outlet-Specific Disclosures 99
3.3 Topic Analysis 100
3.4 Phrase Co-occurrence Network Analysis 101

3.5 Cross-Outlet Recurrence Across Time and Firms 103
4. Findings 105
4.1 S&E Content of Overlapping and Outlet-Specific Disclosures 105
4.2 Accountability Logics Underlying Overlapping and AIF-Specific Disclosures 110
4.3 Temporal and Firm-Level Variation in Cross-Outlet Recurrence 114
5. Discussion 118
Chapter 5: Conclusion 123
References 129
Appendices 146
Appendix A. Variable Definitions 146
Appendix B. Social Keyword Dictionary and Exclusion Criteria 148
Appendix C. Robustness to Cosine Similarity Threshold 153
Appendix D. Institutional Salience Reference Corpus 154
Appendix E. Variable Definitions 157
Appendix F. Representative Sentences by Topic and Lifecycle Stage 158
Appendix G. S&E Dictionary 161
Appendix H. Pooled BERTopic Analysis 166
Appendix I. Variable Definitions 168
Appendix J. OLS Regression Results 169

List of Tables

Table 2.1 Sample Description 17
Table 2.2 Labels and FREX Words for 19 Topics 24
Table 2.3 Labels and Representative Sentences for 19 Topics 26
Table 2.4 Descriptive Statistics for Variables in Regressions 31
Table 2.5 LASSO-Based Estimates of the Relationship between ROA Increase and Topic Adoption Variables 33
Table 2.6 Descriptive Statistics for Filing-Level Tests 38
Table 2.7 LASSO-Based Estimates of the Relationship between ROA Increase and Topic Proportion Variables 40
Table 2.8 LASSO-Based Estimates of the Relationship between ROA Increase and Abnormal Topic Disclosure Variables 42
Table 2.9 LASSO-Based Estimates of the Relationship between CEO Pay Increase and Topic Adoption Variables 44
Table 2.10 LASSO-Based Estimates of the Relationship between Earnings Management and Topic Adoption Variables 46
Table 2.11 LASSO-Based Estimates of the Relationship between Abnormal Returns and Topic Proportion Variables 48
Table 2.12 LASSO-Based Estimates of the Relationship between Topic Adoption Variables and Future Accounting Number Change 49
Table 3.1 Sample Selection 64
Table 3.2 Top 5 Social Topics 72
Table 3.3 Lifecycle Profiles of Dominant Voluntary Social Topics 73
Table 3.4 Descriptive Statistics 75
Table 3.5 Post-LASSO OLS Regression Results: Topic Lifecycle Profiles 77
Table 3.6 Post-LASSO OLS Regression Results: Alignment with Institutional Expectations as Moderator of Topic Lifecycle 79
Table 4.1 Topic Structure of Three Disclosure Categories 105
Table 4.2 S&E Content of Three Disclosure Categories 106
Table 4.3 Network Statistics 111
Table 4.4 Top Phrase Connections in Three Disclosure Categories 111
Table 4.5 Descriptive Statistics 116
Table 4.6 Fractional Logistic Regression Results 117

List of Figures

Figure 2.1 Diagnostic Values by Number of Topics 20
Figure 2.2 Topic Prevalence 29
Figure 4.1 Average Recurrence Ratio over Time 115

Chapter 1: Introduction

Corporate ESG disclosures are not neutral transmissions of information. They are representational practices through which organizational activities and consequences are rendered into particular forms of account, making some things visible while leaving others absent (Burchell et al., 1980; Hopwood, 1987). A substantial body of accounting research has examined why firms disclose and how much they disclose, documenting the legitimacy pressures, institutional expectations, and impression management incentives that shape disclosure behavior (Neu et al., 1998; Deegan et al., 2002; Cho and Patten, 2007; Nazari et al., 2015). Yet this literature has remained comparatively less attentive to three dimensions of ESG disclosure that bear directly on how narrative construction works in practice: the substantive topics managers choose to foreground within their narratives, how specific content is introduced, sustained, and abandoned across successive reporting cycles, and how accounts are positioned and reproduced across the connected documents that together constitute a firm's reporting system. This dissertation takes up each of these dimensions in turn through three empirical studies, each situated in a regulated disclosure setting in which firms are required to file but retain substantial discretion over what to say and how to say it.

Chapter 2 examines topical content in Compensation Discussion and Analysis (CD&A) disclosures. CD&A disclosures are mandatory but narratively discretionary: while the SEC requires firms to explain their executive compensation decisions, managers retain considerable latitude in framing those explanations. Prior research on CD&A has focused largely on stylistic features such as tone and readability (Laksmana et al., 2012; Balsam et al., 2016), leaving the substantive topics managers choose to emphasize and how those choices shift with organizational

circumstances comparatively underexplored. Drawing on impression management theory (Aerts, 2005; Neu et al., 1998) and applying structural topic modeling and LASSO regression to 1,980,034 sentences from 13,057 CD&A filings by S&P 1500 firms over fiscal years 2006 to 2015, I identify 19 topics organized into eight thematic categories and find that managers most frequently foreground topics related to accounting, aligning interests, and pay-for-performance. More importantly, increases in ROA are associated with greater emphasis on aligning interests and compliance and less emphasis on accounting, pay-for-performance, peers, independence, and risk management topics. Additional analyses examine topic patterns associated with declining ROA, rising CEO pay, and earnings management, as well as associations with subsequent market reactions and future accounting performance. Collectively, the findings are consistent with topic selection representing a content-based dimension of impression management that has received less attention in prior research focused primarily on stylistic features such as tone and readability.

Chapter 3 investigates how voluntary social disclosures in Canadian management information circulars (MICs) evolve over time. Prior social disclosure research has documented changes in disclosure volume, tone, and specificity in response to legitimacy pressures and stakeholder expectations (Patten, 1991; Gray et al., 1995). Yet this literature has provided limited insight into whether and how specific pieces of disclosure content are continued or discontinued across successive reporting periods, a distinction that matters because continuity and discontinuity in disclosure content may reveal different aspects of firms' narrative strategies than aggregate changes in volume or style alone. Drawing on 80,890 voluntary social sentences from 12,774 MIC filings by 2,408 Canadian firms spanning 1997 to 2023, I apply FinBERT-ESG embeddings and BERTopic to identify five dominant social topics and develop a sentence-level lifecycle framework to classify each sentence as introduced, carried forward, or abandoned. The findings

show that lifecycle stages differ systematically in their substantive content: introduced disclosure tends to reflect new developments and organizational responses, carried-forward disclosure centers more on established systems and commitments, and abandoned disclosure is more often episodic or time-bound. Filing-level analyses further show that disclosure change varies with topic composition, with greater emphasis on Labour Relations and Workplace Health and Safety associated with higher overall turnover. Greater institutional alignment is associated with greater disclosure continuity overall, although this relationship varies with the relative prominence of selected social topics.

Chapter 4 tests how social and environmental (S&E) accounts are distributed across connected corporate reporting outlets. Prior S&E disclosure research has demonstrated that corporate reports are selective representations of organizational conduct rather than comprehensive accounts of social and environmental performance (Gray, 2007; Michelon et al., 2015). Yet this literature has typically examined individual reports or compared disclosure attributes across firms and time periods, providing limited insight into how related accounts are organized across multiple documents filed within the same reporting cycle. Drawing on 3,416 matched Annual Information Form (AIF) and Management Discussion and Analysis (MD&A) filing pairs from Canadian mining firms between 1997 and 2024, I use FinBERT-ESG sentence embeddings and cosine similarity to classify S&E sentences into overlapping, AIF-specific, and MD&A-specific disclosures, and apply topic modeling, phrase co-occurrence network analysis, and fractional logistic regression to examine how these documentary positions differ in content, accountability logics, and organizational characteristics. The findings show that overlapping disclosures concentrate in formalized accounts of environmental assessment, closure obligations, greenhouse gas regulation, and workplace safety and are organized through a regulatory-

compliance logic. AIF-specific disclosures engage more situated concerns, including project-level impact assessment, community relations, Indigenous land rights, and emissions targets, and are organized through a rights-and-responsibility logic. MD&A-specific disclosures focus more on implementation progress, operational outcomes, and managerial estimates and are organized through a performance-and-estimation logic. Cross-outlet recurrence has increased over time and is positively associated with MD&A S&E disclosure volume while negatively associated with firm size, AIF filing length, and asset intensity. These findings suggest that disclosure selectivity operates not only through what firms disclose but also through how particular accounts are positioned and reproduced across connected reporting documents.

Taken together, the three studies make three broader contributions to accounting research. First, the dissertation contributes to research on disclosure visibility by showing that visibility is not adequately captured by whether information appears in a report or by how much attention it receives at a particular point in time. Across the three studies, visibility is organized along several related dimensions. Chapter 2 shows that firms vary in the substantive issues they foreground within disclosure. Chapter 3 shows that visible content also differs in temporal durability, as disclosure may be newly introduced, carried forward across reporting periods, or subsequently abandoned. Chapter 4 further shows that visibility has a documentary dimension, as related accounts may recur across connected reporting outlets or remain confined to a particular filing. Taken together, these findings extend the analysis of disclosure visibility beyond presence or prominence to encompass substantive emphasis, temporal durability, and documentary positioning. Corporate disclosure is therefore selective not only in what becomes visible, but also in what remains visible and where that visibility is maintained.

Second, the dissertation contributes to research on accountability by showing that the organization of disclosure visibility is consequential for how organizational activities, relationships, and responsibilities are rendered into accounts. Making an issue visible does not necessarily produce a uniform form of account-giving. This is particularly evident in Chapter 4, where recurring and outlet-specific S&E disclosures are organized around different accountability logics, including regulatory compliance, rights and responsibilities, and managerial performance and estimation. Chapters 2 and 3 complement this perspective by showing that firms exercise discretion over which substantive issues receive emphasis and which disclosure content remains part of the reporting narrative over time. Together, these findings suggest that accountability is mediated through the organization of disclosure itself: substantive emphasis, temporal continuity, and documentary positioning organize which organizational matters, obligations, and relationships are made available for external scrutiny. The dissertation therefore extends the analysis of corporate accountability from the provision of information to the ways in which different forms of account-giving are organized and sustained through disclosure.

Third, the dissertation makes a methodological contribution by demonstrating how complementary computational textual techniques can be combined with interpretive analysis to examine complex disclosure processes at scale. Across the three studies, structural topic modeling and BERTopic are used to identify thematic structure; domain-specific language models and semantic similarity measures trace related content across reporting periods and documents; LASSO facilitates the analysis of systematic relationships involving high-dimensional textual measures; and phrase co-occurrence networks identify recurring associations through which disclosure accounts are organized. The methodological contribution lies not in any single technique, but in assembling complementary computational tools around distinct theoretical

questions and combining their outputs with interpretive analysis of representative sentences, topic meanings, lifecycle content, and accountability relations. This approach makes it possible to identify patterned textual variation across large disclosure corpora while retaining attention to the substantive meaning of the patterns detected. It thereby demonstrates how computational text analysis can extend the scale and analytical reach of disclosure research without treating automated classifications as substitutes for theoretically informed interpretation.

The dissertation is subject to several limitations that define the scope of its findings. The three studies examine disclosure within specific regulatory and documentary settings, including U.S. CD&A disclosures, Canadian MICs, and AIF–MD&A filings by Canadian mining firms. The findings may therefore not generalize uniformly across other industries, jurisdictions, or reporting environments. In addition, the textual analyses capture how organizational activities and relationships are represented in disclosure rather than the underlying practices themselves.

The empirical analyses are also primarily associational rather than causal, and the computational methods necessarily involve choices concerning topic identification, semantic similarity, classification thresholds, and network construction. Although these choices are supported by robustness checks and close reading of the underlying text, the findings should be interpreted as theoretically informed analyses of systematic disclosure patterns rather than exhaustive or causal representations of corporate reporting.

Taken together, the three studies approach corporate disclosure as an organized representational practice in which visibility and accountability are shaped through choices about what is emphasized, what is sustained over time, and where particular accounts are positioned. By examining these choices across different regulated reporting settings, the dissertation seeks to provide a more comprehensive understanding of how corporate narratives are constructed and how

they make organizational activities, relationships, and responsibilities available for scrutiny. The following three chapters develop these questions in their respective empirical settings, followed by a concluding chapter that draws together the broader implications of the dissertation.

Chapter 2: Accounting Numbers and Executive Compensation Narratives

Abstract

Compensation Discussion and Analysis (CD&A) disclosures have become increasingly significant not only for investors and senior managers but also for those interested in tracking excessive compensation. Drawing on the impression management theory and applying latent Dirichlet allocation (LDA) topic modeling to a corpus of over 1.98 million sentences from 13,057 CD&A disclosures over the 2006-2015 period, I identify 19 topics across eight major categories. I find that managers most frequently emphasize topics related to the *accounting*, *aligning interests*, and *pay-for-performance*. Further, increases in return on assets (ROA) are associated with greater emphasis on topics related to *aligning interests* and *compliance* while lessening emphasis on topics related to *accounting, pay-for-performance, peers, independence,* and *risk management*. Additional analyses examine topic patterns associated with decreasing ROA, rising CEO pay, and earnings management, as well as links to subsequent market reactions and future accounting performance. Overall, the results are consistent with substantive topic selection representing a narrative dimension of impression management in CD&A disclosures.

Keywords

Accounting performance; CD&A disclosures; impression management; topic modeling

1. Introduction

Executive compensation has faced increasing scrutiny from regulators, investors, and other stakeholders, particularly following high-profile corporate scandals such as Enron and WorldCom. In response, the U.S. Securities and Exchange Commission (SEC) introduced the Compensation Discussion and Analysis (CD&A) section in 2006, requiring firms to disclose their executive pay policies and rationales. While these disclosures follow structured regulatory guidelines, managers retain considerable discretion in framing their narratives, giving rise to strategic communication that influences perceptions of pay fairness and governance (Kim et al., 2025; Laksmana et al., 2012).

Unlike other corporate disclosures, CD&A narratives are explicitly designed to justify executive compensation. Prior research has shown that textual characteristics of CD&A disclosures, such as tone and readability, are affected by executive compensation (Laksmana et al., 2012; Hemmings et al., 2020), and in turn influence outcomes such as say-on-pay voting (Balsam et al., 2016; Wang et al., 2020) and stock performance (Martin et al., 2021). This stream of research draws attention to narrative dimensions of CD&A disclosures, but less is known about narrative content—specifically, the substantive topics companies use to justify executive pay.

To address this research gap, I draw on impression management theory and examine two interrelated research questions. First, *what topics does management choose to emphasize in their CD&A disclosures* (RQ1)? CD&A narratives offer managers flexibility in what and how information is presented. Identifying the underlying thematic content reveals the strategic language managers adopt when legitimizing compensation practices and shaping shareholder perceptions. Second, *what is the relationship between changes in accounting performance numbers and CD&A topic emphasis* (RQ2)? Impression management theory suggests that corporate narratives are

responsive to evolving circumstances (Aerts, 2005; Neu et al., 1998). When financial performance changes, managers may adjust the emphasis of their disclosures to preserve credibility and shareholder support. Together, these questions allow me to unpack both the structure and flexibility of CD&A messaging, moving beyond aggregated textual features to understand how CD&A narratives are constructed.

To answer my research questions, I analyze 1,980,034 sentences from 13,057 CD&A disclosures by 1,437 firms over fiscal years 2006-2015. For RQ1, I first define an *accounting* topic that contains terms such as income, earnings, and taxes. I then use a structural topic modeling algorithm (Roberts et al., 2014), an extension of the latent Dirichlet allocation (LDA) model developed by Blei et al. (2003), to identify 18 additional conceptually significant topics in CD&A disclosures. The topics cover content such as long-term incentives, bonus calculation mechanisms, compensation survey sources, and independent compensation advisors. I further classify the 19 individual topics into eight broader thematic categories: *accounting, aligning interests, pay-for-performance, peers, independence, basic introduction, risk management,* and *compliance*. I find that managers devote greater narrative attention to topics in the *accounting, aligning interests*, and *pay-for-performance* categories. For RQ2, I use the least absolute shrinkage and selection operator (LASSO) regression technique (Tibshirani, 1996), regressing the year-over-year increase in return on assets (ROA) on both the 19 individual topics and the eight combined topic variables. Collectively, I find that increases in ROA are associated with greater emphasis on topics related to *aligning interests, basic introduction,* and *compliance* and lower emphasis on topics related to *accounting, pay-for-performance, peers, independence,* and *risk management*.

Robustness tests examining "abnormal" changes in topic choice further suggest systematic associations between changes in accounting performance and the emphasis placed on selected

topics, particularly those related to *aligning interests* and *compliance*. These results are robust to a variety of sensitivity analyses. Additional analyses also examine topic choices associated with decreases in ROA and increases in CEO pay, as well as associations between topic choices and subsequent market reactions and accounting performance.

This study makes three central contributions to the accounting literature. First, it enhances research on compensation disclosures by offering a detailed analysis of CD&A narratives. While previous studies have explored CD&A disclosures through textual features such as readability (Laksmana et al., 2012) and tone (Balsam et al., 2016), the thematic structure and information content within these disclosures remain underdeveloped areas of research. My study extends this body of work by revealing how CD&A disclosures are built around specific topics, providing a deeper understanding of CD&A disclosures. Moreover, save for a study by Packard et al. (2023), prior studies typically quantify the textual attributes based on the entire CD&A section, which potentially overlooks crucial information at the intra-document level. This enables a richer understanding of the micro-level structures managers use to legitimize executive pay, thereby addressing a significant gap in the compensation disclosure literature.

Second, this study advances the accounting literature on impression management by identifying substantive topic selection as an additional narrative dimension associated with performance fluctuations. Prior research adopting the impression management lens has often examined changes in textual characteristics such as tone, readability, and length (e.g., Aerts, 2005; Neu et al., 1998; Leung et al., 2015) to infer how managers influence stakeholder perceptions, for example by framing performance more favorably or adjusting disclosure length. However, these studies often overlook how firms manage impressions by choosing *what* to talk about—namely, the specific compensation-related topics they highlight or downplay. By showing that the

prevalence of specific topics varies systematically with accounting performance, this study extends impression management research from stylistic characteristics to the substantive content of corporate narratives.

Third, this study contributes to accounting research using topic modeling by integrating the structural topic model with LASSO regression and a range of complementary analyses. Through various additional tests, I provide insights into how other management actions, such as CEO pay increases and earnings management, are associated with the selection of topics in CD&A disclosures. Additionally, I expand the scope of topic analysis by examining market reactions to specific CD&A topics. Overall, the tests establish a robust foundation for future research on corporate communication strategies and their broader implications.

The rest of the paper is structured as follows. Section 2 introduces prior literature on CD&A disclosures while Section 3 develops the theoretical framework and research questions. Section 4 describes the data and method. Sections 5 and 6 present the main findings and additional analyses. Section 7 summarizes the study's findings and limitations along with its implications for future research.

2. Background

CD&A disclosures serve as a platform for top management to explain decisions on executive compensation (Bloomfield, 2021). Introduced by the SEC in 2006, the CD&A section is intended to provide investors with insights into how compensation policies align with firm performance. While the disclosures are guided by a formal regulatory structure, managers retain considerable discretion in how they frame and emphasize their explanations. This flexibility allows managers to tailor the narratives to organizational contexts and shareholders' expectations. The importance of CD&A disclosures has grown substantially following the Dodd-Frank Act 2010,

which introduces a mandatory "say-on-pay" policy, granting shareholders of U.S. public firms the right to vote on executive compensation (Packard et al., 2023).

A growing literature has analyzed CD&A disclosures by focusing on their textual and stylistic attributes (e.g., Balsam et al., 2016; Hemmings et al., 2020; Hooghiemstra et al., 2017; Laksmana et al., 2012; Martin et al., 2021; Wang et al., 2020). For example, Balsam et al. (2016) find managers tend to use a positive tone in CD&A to reduce voting dissent among shareholders. Laksmana et al. (2012) find that firms with overpaid CEOs tend to issue disclosures with lower readability, suggesting a strategic use of complexity to obscure justification. Martin et al. (2021) document that changes in a firm's CD&A disclosures over time are predictive of the firm's future stock returns. Collectively, these studies reveal that CD&A narratives are not merely compliance exercises but carry significant informational value for investors and governance outcomes.

Complementing this linguistic focus, a growing body of research has begun to examine the actual content of CD&A disclosures (Packard et al., 2023; LaViers et al., 2024; Kim et al., 2025). This line of work investigates whether and how firms explain specific elements of executive compensation narratives. For example, Packard et al. (2023) analyze performance-related content and find that firms tend to increase performance-based justification when executive pay is high relative to performance outcomes. Similarly, LaViers et al. (2024) suggest that disclosures about the CEO pay ratio become more prominent as the ratio increases, implying a strategic response to potentially sensitive governance metrics. These studies provide valuable insight into how managers selectively frame sensitive compensation metrics. However, this line of research often isolates specific disclosure elements or focuses on narrowly defined indicators, offering only a partial view of the broad narrative strategies of the entire CD&A disclosures.

Taken together, prior research has established that CD&A disclosures are strategically constructed and consequential for investor reactions. Yet existing studies often analyze either the overall style or isolated content fragments, rather than examining how the disclosure as a whole is composed and how different parts are prioritized. Addressing this gap is particularly important in light of increasing demand for compensation transparency and the rising complexity of executive pay design.

3. Theoretical Framework

3.1 Topics in CD&A Disclosures

This study draws on impression management theory to better understand how managers craft CD&A narratives. Rooted in social psychology (Schlenker et al., 1980), impression management refers to the purposeful construction of messages intended to control how an audience perceives the speaker. This behavior is particularly relevant in corporate reporting, where managers are required to justify their actions and decisions to external stakeholders (Aerts, 2005; Godfrey et al., 2003). Because stakeholders lack direct visibility into managerial processes, they rely heavily on corporate disclosures to interpret organizational performance. This information asymmetry creates managerial incentives to selectively present information that enhances perceptions of firms or minimizes the impact of negative outcomes.

A broad body of accounting literature (Abeysekera, 2014; Aerts and Yan, 2017; Bozzolan et al., 2015; Cooper and Slack, 2015; Leung et al., 2015) has applied impression management theory to interpret corporate disclosures such as annual reports, press releases, and social media announcements, highlighting how firms strategically employ narrative characteristics such as length and tone to influence stakeholder understanding. For example, Leung et al. (2015) find that narrative minimalism, or the practice of disclosing less information, is frequently used in

operational reviews of annual reports to obscure unfavorable developments. Bozzolan et al. (2015) demonstrate that firms adapt the tone of their disclosures depending on the perceived importance of different stakeholder audiences. Collectively, these studies reveal an effort by firms to control both the framing and prominence of disclosed content in pursuit of favorable stakeholder evaluations.

Among various forms of corporate narrative, CD&A disclosures stand out as a particularly strategic arena for impression management given the sensitivity of executive compensation among stakeholders. Although guided by regulatory requirements, these disclosures grant managers significant discretion in how compensation decisions are justified. This narrative flexibility allows managers to highlight themes that reinforce the perceived fairness and alignment of executive pay, while selectively downplaying elements that may trigger shareholder concern or criticism. Building on this rationale, I pose the following research question:

RQ1: What topics does management choose to emphasize in their CD&A disclosures?

3.2 Changes in Accounting Performance and Choice of CD&A Topics

Impression management theory holds that corporate narratives typically function in a reactive mode, especially under conditions of changing performance (Aerts, 2005; Cho et al., 2010; Elshandidy & Kamel, 2024; Moreno et al., 2019; Neu et al., 1998). For example, Moreno et al. (2019) document that firms increase the use of vague, tentative language in annual reports during poor performance periods, suggesting an effort to blur responsibility or causality. Similarly, Cho et al. (2010) find that environmentally poor-performing firms adopt a more optimistic yet uncertain tone in their environmental 10-K disclosures. These studies collectively indicate that disclosure features vary systematically with performance, underscoring their role in impression management.

Extending this logic to CD&A disclosures, I posit that thematic content in CD&A disclosures is likely to be associated with changes in accounting performance. When accounting performance improves, topics such as pay-for-performance, aligning interests, and compliance may receive greater emphasis, foregrounding the appropriateness of executive rewards by connecting compensation to firm performance, shareholder interests, and established governance practices. Conversely, weaker performance may coincide with lower emphasis on themes such as accountability or risk management and greater emphasis on other aspects of compensation disclosure. More broadly, the substantive composition of CD&A narratives may vary systematically with accounting performance. Consequently, while I do not know *a priori* which topics managers might emphasize (see RQ1), I expect topic emphasis to differ across changes in accounting performance. This leads to my second research question:

RQ2: What is the relationship between changes in accounting performance numbers and CD&A topic emphasis?

4. Data and Method

4.1 Sample Selection

I begin with 13,433 proxy statements filed by 1,474 S&P 1500 firms over fiscal years 2006 to 2015,[1] downloaded from the SEC's EDGAR database. Using custom Python code, I extract CD&A disclosures from these filings by identifying section delimiters such as "compensation discussion & analysis" and "compensation committee report," followed by manual verification for

[1] The data end in 2015 on account of the extensive work involved in reliably parsing the CD&A data. Given the effort required, I believe 10 years' data is more than sufficient to test the relationship between accounting number changes and topic emphasis. Moreover, during the sample period, there are no substantial social or other exogenous changes (e.g., pandemic) that may have affected management's impression management incentives and disclosing behavior. For regulatory change, it is important to note that the introduction of the "say-on-pay" vote in 2011 may have influenced CD&A disclosures. However, I view this as an evolution of existing practices rather than a disruptive shift (Packard et al., 2023). Therefore, I believe the findings using a dataset from 2006 to 2015 capture broader trends in disclosure behaviors.

accuracy. I then clean the extracted text by removing HTML code tags, hidden javascript code, table formatting, images, and page numbers to ensure an analyzable dataset.

After removing 15 duplicate CD&A filings and 37 firms lacking Compustat data (accounting for 361 firm-year observations), I obtain a sample of 13,057 CD&A documents from 1,437 firms. Subsequently, following previous LDA-based accounting studies (e.g., Neu & Saxton, 2022) and Packard et al.'s (2023) sentence-level study on CD&A disclosures, I parse the 13,057 CD&A documents into 2,258,574 sentences. I exclude 278,540 sentences shorter than 30 characters to ensure semantic relevance, yielding a final sample of 1,980,034 sentences across 13,057 documents. Panel A of Table 2.1 summarizes the sample construction process. Panel B of Table 2.1, meanwhile, reports the yearly distribution of 13,057 CD&A disclosures, showing a balanced sample across years. Panel C of Table 2.1 then reports the industry composition of my 1,437 sample firms and shows that the composition of the sample is similar to that of Compustat firms.

Table 2.1 Sample Description

Panel A Sample selection

	Number of CD&A Filings	Number of CD&A sentence-level observations
All CD&A documents filed by S&P 1500 firms over fiscal years 2006 to 2015	13,433	
Less: duplicated filings	(15)	
Less: Unavailable firm-year observations in Compustat database	(361)	
Primary sample	13,057	2,258,574
Less: sentences with less than 30 characters		(278,540)
Final sample		1,980,034

Note: Panel A presents sample selection details. The final sample consists of 1,980,034 sentences from 13,057 CD&A disclosures over fiscal years 2006-2015.

Panel B Year distribution of sample CD&A disclosures		
Fiscal year	Number of CD&As	Proportion
2006	1,035	7.93%
2007	1,251	9.58%
2008	1,282	9.82%
2009	1,295	9.92%
2010	1,322	10.12%
2011	1,332	10.20%
2012	1,354	10.37%
2013	1,378	10.55%
2014	1,395	10.68%
2015	1,413	10.82%
Total	*13,057*	

Note: Panel B presents the number and proportion of CD&A disclosures in each fiscal year.

Panel C Industry composition of sample firms			
Standard Industrial Classification (code)	Number of sample firms	Proportion of sample firms	Proportion of firms in Compustat database
Agriculture, Forestry, And Fishing (0100-0999)	3	0.21%	0.39%
Mining (1000-1499)	58	4.04%	10.23%
Construction (1500-1799)	23	1.60%	0.73%
Manufacturing (2000-3999)	552	38.41%	34.03%
Transportation, Communications, Electric, Gas, And Sanitary Services (4000-4999)	136	9.46%	7.65%
Wholesale Trade (5000-5199)	49	3.41%	2.18%
Retail Trade (5200-5999)	109	7.59%	4.06%
Finance, Insurance, And Real Estate (6000-6799)	306	21.29%	21.37%
Services (7000-8999)	198	13.78%	17.01%
Public Administration (9100-9999)	3	0.21%	2.34%
Total	*1,437*		

Note: Panel C presents the industry composition of the 1,437 sample firms. The panel shows that the industry composition of my sample is similar to that of the Compustat firms.

4.2 Measuring CD&A Topics

I identify topics in the CD&A disclosures through a two-stage topic process where I first deductively pre-define an "accounting" topic and then employ an LDA model (Blei et al., 2003) to discover all other meaningful topics. In the first stage, I begin by generating a document-term matrix that allows me to identify a frequency listing of all the terms in the sample of 1,980,034 CD&A sentences. From the top 3,000 terms in the frequency listing, I then identify a list of relevant

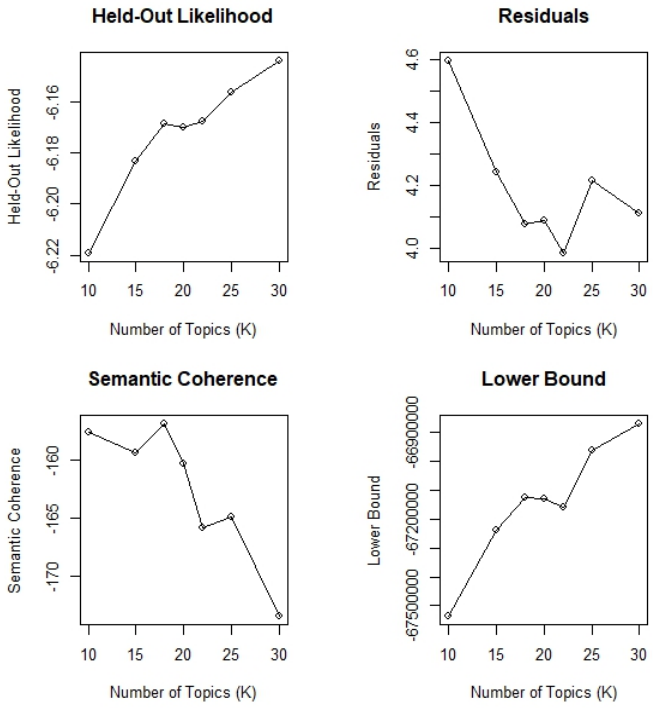
accounting terms: *tax, taxes, cost, costs, expense, expenses, income, earnings, revenue, revenues, netincome,*[2] *profit, asset, assets, ebit, ebitda*, and *accounting*, which I use to identify the "accounting" topic ($TOPIC_{Accting}$). This pre-identification of the accounting topic is predicated on the critical role that accounting terms have been shown to play in framing key financial and performance-related results in corporate disclosures. For example, Aerts (1994) points out that managers select "technical accounting terms," such as profit and loss accounts, to explain negative performance in accounting narratives.

In the second stage, I remove accounting terms and stopwords (e.g., compensation, the, and, etc.) from the document-term matrix and utilize LDA to identify other latent topics within CD&A sentences. LDA is a Bayesian probabilistic approach used to uncover latent connections between words across documents and group them based on similar topics (DiMaggio et al., 2013).[3] The approach enables researchers to identify patterns within a large number of documents that would otherwise be difficult to manually read and analyze.[4] More specifically, I employ the structural topic model (STM) algorithms developed by Roberts et al. (2014). The *STM* package in *R* (Roberts et al., 2019) implements a Bayesian probabilistic LDA model along with key diagnostics that facilitate the identification of the optimal number of topics, or what is known as a *k* value. While it is ultimately up to the researcher to decide on the *k* value that best conforms with the data and the particular study context, these quantitative diagnostics help identify the most appropriate range of topics (Bao & Datta, 2014; Bellstam et al., 2021). Figure 2.1 summarizes the

[2] The term "net income" is replaced by "netincome" before running text processing.
[3] As noted by DiMaggio et al. (2013), LDA treats documents as a "bag of words" and detects co-occurrence of words in the documents to estimate the topic distribution of each document.
[4] Prior studies (Bao & Datta, 2014; Dyer et al., 2017; Hoberg & Lewis, 2017) have applied LDA to corporate disclosures. Bao and Datta (2014), for example, used the LDA approach to identify 30 risk factor topics in 10-K reports from 2006-2010. Similarly, Dyer et al. (2017) analyzed trends in the topical content in 10-K reports. Collectively, these studies demonstrate the effectiveness of the LDA technique in identifying themes within accounting disclosures. I believe this approach could significantly enhance the understanding of CD&A disclosures as well; by uncovering how senior managers emphasize and de-emphasize different topics to justify compensation practices, this approach could provide deeper insights into the strategic messaging efforts of top management.

key quantitative diagnostics, which suggest that the optimal number of topics is somewhere between 10 and 30. Based on these diagnostics, I identify that the optimal number of topics in the document corpus is 18. To address RQ1, I use the *STM* package to generate representative sentences for each topic. I then assign a descriptive label to each topic based on my interpretation of both the FREX words and the representative sentences.

Figure 2.1 Diagnostic Values by Number of Topics



Note: Figure 2.1 presents diagnostic values in terms of Held-Out Likelihood, Semantic Coherence, Lower Bound, and Residuals for a range of topic numbers from k = 10 to k = 30. The optimum number of topics should maximize Held-Out Likelihood, Semantic Coherence, and Lower Bound while minimizing Residuals. The topic model solutions indicate that the topic model does the best trade-off among different diagnostic indicators when the topic number is 18. The tests are performed based on 1,980,034 sentences from 13,057 CD&As over fiscal years 2006 to 2015.

To address RQ2, I construct a set of binary variables that capture whether each CD&A sentence includes a given topic. For this I follow prior research (e.g., Roberts et al., 2019; Neu et al., 2022) by relying on the *frequent-and-exclusive* (FREX) words that are generated for each topic by the STM algorithms and serve to distinguish one topic from another. For each of the 19 topics,

I define an indicator variable, $TOPIC_i$, that equals 1 if a sentence contains at least one FREX word associated with topic *i*, and 0 otherwise. In addition to analyzing individual topics, I also group them into eight broader thematic categories based on their content coherence. For these combined groupings, I define $TOPIC_i$ as 1 if any FREX word from the combined topic *i* appears in a sentence, and 0 otherwise. This structure enables both granular and aggregated analysis of how topic emphasis varies with accounting performance changes.

4.3 Measure of Changes in Accounting Performance Numbers

My proxy for the change in accounting performance is the year-over-year change in ROA (net income over total assets). When crafting the CD&A, managers already have access to the current year's financial data – or at least a reliable estimate. Consequently, it is a change in the *current* fiscal year's accounting numbers, especially changes in key performance indicators, that is most relevant to examining variation in managers' subsequent narrative choices in the CD&A. To top management, ROA is significant because it reflects senior managers' performance and is closely related to their personal wealth (Merkley, 2014). Moreover, my adoption of ROA as a proxy for accounting performance is consistent with prior textual analysis research (e.g., Matsumoto et al., 2011; Packard et al., 2023) that argues ROA offers a standardized means of comparing companies across different industries and years, neutralizing the effects of size. The main analyses focus on the year-over-year *increase* in ROA.[5] I define *ROA_INCREASE* as an indicator variable that equals one if ROA for the current fiscal year increases, and zero otherwise.[6]

[5] Additional analyses examine topic patterns associated with a decrease in ROA.

[6] To ensure the robustness of the results, I also create a version of the variable that equals one if there is a 2.5% or greater increase in ROA compared to the previous year, otherwise zero. The untabulated results for this variable are consistent with the results using *ROA_INCREASE*, suggesting no substantial heterogeneity in the associations between different magnitudes of ROA increases and topic adoption.

4.4 Control Variables

I also include a series of control variables found to be relevant in the prior literature. First, I use Compustat data to generate six measures related to firm fundamentals: *ASSETS*, the natural logarithm of total assets (Laksmana et al., 2012); *LEVERAGE,* total debt divided by total assets (Lang & Stice-Lawrence, 2015); *TURNOVER,* the ratio of total revenue to total assets, given that revenue is related to both ROA and executives' revenue-based pay; *MTB,* the ratio of market value to book value of total assets (Elshandidy & Kamel, 2024); *PPE,* net property, plant, and equipment divided by total assets (Kim et al., 2023); and *CAPEX,* capital expenditure divided by total assets (Vo & Canil, 2019). I then use Execucomp data to control for total executive compensation, where *EXECPAY* is calculated as the ratio of total executive compensation to total revenue, multiplied by 100 (Hooghiemstra et al., 2017). In addition, I control for CD&A *LENGTH*, measured as the word count scaled by 1,000. For all of the above controls, I replace missing values with their average values and winsorize at the 1% and 99% levels. Lastly, I include year fixed effects to control for time-specific factors and industry fixed effects to control for potential industry-specific differences. Variable definitions are presented in Appendix A.

4.5 LASSO Regression Model

To evaluate the relationship between accounting number change and topical content for RQ2, I use the LASSO regression technique developed by Tibshirani (1996) and which has become more popular in accounting research (Basu et al., 2022; Neu & Saxton, 2023). LASSO introduces a penalty term that shrinks less informative coefficients toward zero, thereby simplifying the model structure (Hastie et al., 2023). This regularization process offers several advantages in my context. First, it enables variable selection by identifying relevant CD&A topics (those with non-zero coefficients) that are most strongly associated with changes in ROA. Second,

it mitigates overfitting by constraining model complexity. Third, it addresses multicollinearity, which is common when topic variables overlap or co-occur across disclosures.

It is also important to highlight the temporal ordering underlying the analysis. Although ROA change and CD&A topics are measured within the same fiscal year, the CD&A discussion is prepared after management observes the financial performance being discussed. I therefore examine whether variation in accounting performance is systematically associated with the topics subsequently emphasized in CD&A disclosures. Given the predictive structure of LASSO, the model uses topic variables to predict whether an increase in ROA occurred, allowing the analysis to identify the topics most strongly associated with ROA changes. This use of LASSO is consistent with other topic modeling research (e.g., Morgan-Lopez et al., 2017; Neu & Saxton, 2023; Rao et al., 2010) that seeks to predict a temporally precedent feature (*ROA_INCREASE*, in my case) based on the topics that are included in the document.[7] While LASSO provides a useful approach for identifying the topics most strongly associated with ROA changes, I also ran logistic regressions in which each topic was the dependent variable in a separate regression. These untabulated tests corroborate the LASSO regression results. At the same time, I must be cautious in noting that the tests are chiefly associational and, despite a series of robustness tests and additional analyses, I cannot fully rule out endogeneity concerns.

[7] There is a growing literature that similarly seeks to use such LASSO and other machine learning techniques to predict, for example, stable speaker characteristics such as sex, age, or profession (i.e., temporal precursors) based on textual characteristics in the documents they write (Morgan-Lopez et al., 2017; Rao et al., 2010). Analogous to this prior research, the LASSO machine learning model is used to predict *ROA_INCREASE* based on the topics emphasized in the CD&A.

5. Findings

5.1 CD&A Topics

As described in the methods section, I use a combination of quantitative diagnostics and inductive analysis of the data and the topic model output to determine that the best LDA solution contains 19 discrete topics. Based on my analysis of each of the topics returned by the 19-topic LDA model, I then manually label each of the topics. After reading the content of the topics, I also find that the 19 topics could be meaningfully grouped into eight broader, coherent categories. Table 2.2 shows the labels for each of the 19 topics, organized by the eight broader categories, along with the associated FREX (*frequent-and-exclusive*) words as calculated by the STM algorithms.[8] Table 2.3, meanwhile, shows example sentences for each topic.

Table 2.2 Labels and FREX Words for 19 Topics

19 individual topics and descriptions by topical category	FREX words
Accounting	
$TOPIC_{Accting}$: Accounting terms	tax, taxes, cost, costs, expense, expenses, income, earnings, revenues, net income, profit, asset, assets, ebit, ebitda, accounting
Aligning interests	
$TOPIC_{Incentives}$: Long-term incentives	order, tie, portion, ensur, sinc, retent, remain
$TOPIC_{Voting}$: Shareholders' voting	vote, advisori, saydexi, approxim, propos, consider, express
$TOPIC_{Goals}$: Firms' long-term goals	success, attract, leadership, key, import, strateg, experi
$TOPIC_{Efficiency}$: Short-term incentives to maximize efficiency	mix, sheet, short, invest, effici, improv, balanc
Pay-for-performance	
$TOPIC_{Stock}$: Stock grant calculation	rsus, common, quarter, dividend, close, equal, ltip
$TOPIC_{Non\text{-}GAAP}$: Non-GAAP performance measures calculation	exclud, charg, calcul, dollar, item, loss, amort
$TOPIC_{Bonus}$: Bonus calculation	threshold, maximum, metric, fund, pre, pool, minimum
Peers	
$TOPIC_{Peer}$: Peer firms	presid, vice, inc, senior, resourc, human, corp
$TOPIC_{Market}$: Market survey sources	survey, industri, median, similar, size, comparison, benchmark

[8] The structural topic modeling algorithm provides four measures for topic words. HPROB shows the highest probability words; FREX are the words that are both frequent and exclusive within a topic; LIFT and SCORE are measures of the frequency of word occurrence in the corresponding topic relative to the frequency of occurrence in the rest of the corpus. Following prior research (Roberts et al., 2019; Neu et al. 2022), I use the FREX words to interpret topical content and quantify identified topics.

Independence	
$TOPIC_{Committee}$: Committee's discretion	discret, messr, author, discretionari, final, sole, howev
$TOPIC_{Advisor}$: Independent advisor	independ, member, request, assess, assist, pearl, input
Basic introduction	
$TOPIC_{Elements}$: Compensation package elements	current, packag, histor, talli, potenti, various, vehicl
$TOPIC_{Introduction}$: Introductory sentences	page, see, forth, statement, column, proxi, shown
$TOPIC_{Overview}$: CD&A overview	philosophi, analysi, princip, overview, refer, appli, specif
$TOPIC_{Termination}$: Termination payments and retirement plans	etir, termin, insur, disabl, defer, sever, life
Risk management	
$TOPIC_{Risk}$: Risk management	releas, materi, like, affect, advers, secur, hedg
$TOPIC_{Clawback}$: Clawback policy	previous, restat, due, part, seek, subsequ, basi
Compliance	
$TOPIC_{Compliance}$: Compliance with Internal Revenue Code	code, ownership, guidelin, deduct, limit, adopt, intern

Note: Table 2.2 presents terms for the pre-defined accounting topic and FREX (frequent-and-exclusive) words for the 18 latent topics extracted from 1,980,034 sentences in 13,057 CD&A disclosures over fiscal years 2006 to 2015. The 19 topics are presented according to eight broader topical categories.

I present the 19 topics organized by the eight broader categories. First, there is an *accounting* topic ($TOPIC_{Accting}$), the only topic in the accounting category. As shown in Tables 2.2 and 2.3, this topic involves information about accounting standards and brief descriptions of firm fundamentals such as revenue, profit, and costs, etc.

The second category, *aligning interests,* comprises four different topics about the alignment of interests between executives and shareholders. To start, $TOPIC_{Incentives}$ conveys how compensation designs provide long-term incentives. $TOPIC_{Voting}$ reports the frequency of future say-on-pay votes, implying a significant role for shareholder voting on compensation design. $TOPIC_{Goals}$ discusses how compensation plans motivate executives to contribute to firms' long-term goals. Lastly, $TOPIC_{Efficiency}$ focuses on how short-term incentives are used to drive operational effectiveness.

These disclosures collectively aim to assure stakeholders that executive compensation serves organizational, not merely personal, goals.[9]

Table 2.3 Labels and Representative Sentences for 19 Topics

19 individual topics	Representative sentences
$TOPIC_{Accting}$	Accounting treatment: effective January 1, 2006, we adopted Statement of Financial Accounting Standards No. 123(r), which requires that companies recognize in their financial statements the cost of employee services received in exchange for awards of equity instruments based on the grant date fair value of those awards.
$TOPIC_{Incentives}$	Long-term incentives are designed to tie a significant portion of our named executive officers' compensation to our performance, create a meaningful alignment of our named executive officers' financial interests with those of stockholders, and encourage long term retention.
$TOPIC_{Voting}$	In addition, the compensation committee and the board recommended that future say on pay votes be held annually, and approximately 87% of the votes cast on the say on pay frequency vote proposal were in favor of holding a say on pay vote every year.
$TOPIC_{Goals}$	To achieve our long-term strategic priorities, it is critical that we are able to attract, retain and motivate talented professionals who are committed to these priorities while simultaneously aligning their incentives with the interests of our clients and shareholders.
$TOPIC_{Efficiency}$	Short-term incentive plans: the company has two short-term incentive plans, both of which are cash bonus incentive plans designed to focus executive officers and other participants on maximizing efficiency and generating strong financial performance.
$TOPIC_{Stock}$	Stock options granted by us have an exercise price equal to the fair market value of our common stock on the day of grant, typically vest over a three-year period (with 33% vesting 12 months after the grant date and the remainder vesting ratably each year thereafter based upon continued employment) and generally expire five years after the date of grant.
$TOPIC_{Non\text{-}GAAP}$	EBITDA, as adjusted, is calculated as net income (loss) excluding income (loss) from discontinued operations (net of tax), cumulative effect of accounting changes (net of tax), income taxes, interest expense (including debt extinguishment costs and gain or loss on termination of interest rate swaps), depreciation and amortization expense, impairment charges, restructuring charges, non-cash gains or losses from foreign currency exchange rate changes recorded on intercompany obligations, expensed acquisition costs and other charges.
$TOPIC_{Bonus}$	Under this metric, if the company's FFO growth per share does not exceed any of these companies, no bonus is earned; if the company's FFO growth per share exceeds one of these companies, a bonus of 20% of base salary is earned; if the company's FFO growth per share exceeds two of these companies, a bonus of 40% of base salary is earned; and if the company's FFO growth per share exceeds all three of the peer companies, the maximum bonus of 60% of base salary is earned.
$TOPIC_{Peer}$	As a result, the 2014 peer group consisted of Bill Barrett Corp., Carrizo Oil and Gas, inc., Emerald Oil, inc., Forest Oil Corporation, Goodrich Petroleum Corp., Gulfport Energy Corp., Halcon Resources Corp., Kodiak Oil & Gas Corp., Magnum Hunter Resources Corp., Oasis Petroleum Inc., Resolute Energy Corp., Rsetta Resources Inc., Swift Energy Company, and Triangle Petroleum Corporation.
$TOPIC_{Market}$:	For fiscal year 2011, these survey sources were the Hewitt TCM Executive Survey, the Mercer Benchmark Database Survey, the Towers Perrin CDB Executive Survey, and the Watson Wyatt Top Management Survey.
$TOPIC_{Committee}$	The compensation committee reserves the right under the plan to exercise negative discretion to reduce the payout for any executive officer by up to 25% or to eliminate payouts if it deems appropriate.
$TOPIC_{Advisor}$	As required by SEC and NYSE rules, the committee has assessed the independence of Pearl Meyer and concluded that Pearl Meyer's work did not raise any conflicts of interest.

[9] The findings here corroborate existing research regarding the objectives of incentive payments. For example, Kuang and Qin (2009) discover that performance-vested stock options in compensation contracts aim to align interests, while Richardson and Waegelein (2002) find that long-term incentive plans are effective in enhancing annual returns.

$TOPIC_{Elements}$	In establishing the program for a year, the compensation committee also considers the various elements of our executive compensation packages, including whether to increase salaries in general or for specific executives; whether to change potential payouts as a percentage of salary; and whether to alter the mix between cash and equity compensation.
$TOPIC_{Introduction}$	For a description of the perquisites and the attributed costs of these benefits, see our summary compensation table on page 34 of this proxy statement.
$TOPIC_{Overview}$	Compensation Discussion and Analysis Introduction: this Compensation Discussion and Analysis (CD&A) explains our compensation program design and how our program operated in with respect to our principal executive officer, our principal financial officer, and our three other most highly compensated executive officers, collectively, our named executive officers.
$TOPIC_{Termination}$	We also have four non-qualified supplemental retirement plans: (1) the pension supplemental executive retirement plan, or pension SERP, (2) the pension restoration supplemental executive retirement plan, or pension restoration SERP, (3) the 401(k) restoration supplemental executive retirement plan, or 401(k) restoration SERP, and (4) the 401(k) supplemental executive retirement plan, or 401(k) SERP.
$TOPIC_{Risk}$	Compensation policies relating to risk management the committee believes that our compensation policies and practices do not encourage unnecessary or excessive risk taking and that any risks arising from our compensation policies and practices are not reasonably likely to have a material adverse effect on the corporation
$TOPIC_{Clawback}$	Compensation recoupment (clawback) policy: the recoupment policy requires each executive officer to reimburse the company for all or a portion of any annual or long-term incentive compensation paid to the executive officer based on achievement of financial results that were subsequently the subject of a restatement due to error or misconduct regardless of whether the executive officer was responsible for the error or misconduct so long as no payment or award or a lower payment or award would have been made to the officer based on the restated results.
$TOPIC_{Compliance}$	Compliance with section 162(m) of the Internal Revenue Code of 1986, as amended (the "code") generally provides that compensation paid to certain of the named officers in excess of $1,000,000 cannot be deducted by us for federal income tax purposes, except to the extent such compensation constitutes qualified performance-based compensation.

Note: Table 2.3 presents representative sentences for the 19 topics extracted from 1,980,034 sentences in 13,057 CD&A disclosures over fiscal years 2006 to 2015.

The third category, *pay-for-performance,* includes three topics centered on the mechanisms that link pay to firm performance. $TOPIC_{Stock}$ focuses on the details of how stock options are priced and structured. $TOPIC_{Non\text{-}GAAP}$ highlights the use of non-GAAP performance measures such as EBITDA in determining executive pay. $TOPIC_{Bonus}$ covers how bonuses are determined. These topics reflect an effort to demonstrate transparency and reinforce a strong pay-performance link—a key theme in impression management strategies used to legitimize executive compensation.

The fourth category comprises two different topics employed to rationalize *executive compensation levels* in relation to peer groups and market benchmarks. Specifically, $TOPIC_{Peer}$ lists peer companies that firms discuss to show that they are providing market-level rewards to executives and that compensation designs are aligned with peer group practices. $TOPIC_{Market}$,

meanwhile, discusses the use of external compensation surveys (e.g., the Tower Perrins Survey) to support pay decisions. These disclosures function to legitimate compensation practices by signaling market competitiveness and fairness.

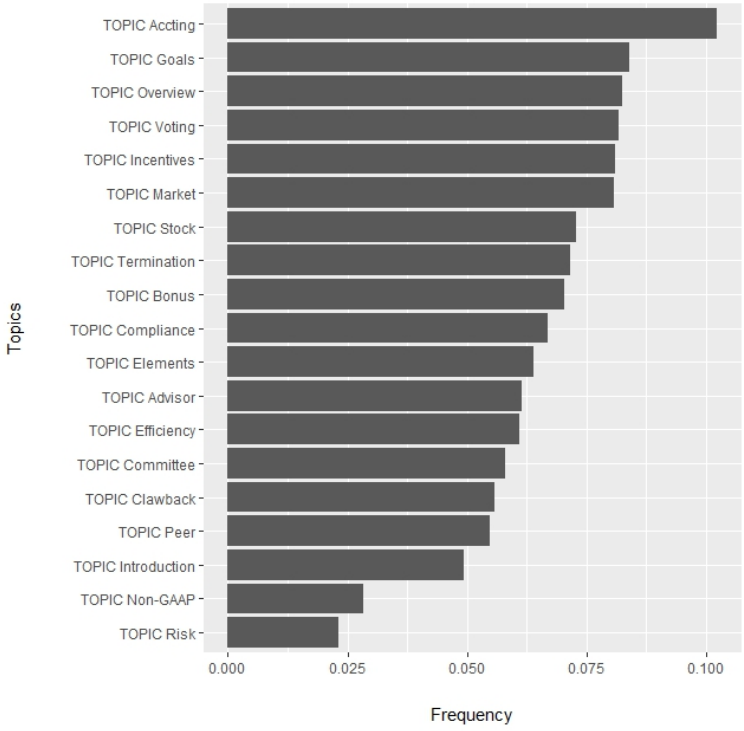
In the fifth category of topics, which I label *independence,* managers highlight the compensation committee's authority to adjust executive compensation ($TOPIC_{Committee}$) and the role of outside advisors ($TOPIC_{Advisor}$). These topics aim to reinforce the procedural integrity of compensation decisions. By emphasizing independence and third-party validation, firms seek to mitigate perceptions of self-serving behavior.

In the sixth category of topics, managers complement the above justifications by presenting a *basic introduction*. $TOPIC_{Elements}$ discusses details related to various components of the compensation packages such as non-monetary payments (e.g., cars, club memberships, etc.). $TOPIC_{Introduction}$ and $TOPIC_{Overview}$ set the stage for the narrative by describing the purpose and structure of the CD&A. $TOPIC_{Termination}$ addresses severance or retirement payments. These disclosures play an informative role and help anchor the justifications discussed elsewhere in the filing.

The seventh category, *risk management,* features two topics. $TOPIC_{Risk}$ discusses compensation plans designed to avoid excessive risk-taking or adverse effects on firms, while $TOPIC_{Clawback}$ points to clawback policies and enforcement to protect the firm's interests.

The last category of topics is *compliance,* which consists of a single topic. $TOPIC_{Compliance}$ discusses how executive compensation designs comply with the Internal Revenue Code, underscoring a commitment to regulatory adherence.

Figure 2.2 Topic Prevalence



Note: Figure 2.2 presents the prevalence ranking of the pre-defined accounting topic and the 18 latent topics extracted from 1,980,034 sentences in 13,057 CD&As over fiscal years 2006 to 2015.

Figure 2.2 shows the frequency of each topic. The accounting topic ($TOPIC_{Accting}$) is the most frequently utilized to structure the CD&A section, appearing in over 10% of all sentences. Three topics in the *aligning interests* category – $TOPIC_{Incentives}$, $TOPIC_{Voting}$, and $TOPIC_{Goals}$ – are also among the top five most-frequent topics and occur in just over 8% of sentences. Two of the topics in the *pay-for-performance* category, $TOPIC_{Stock}$ and $TOPIC_{Bonus}$, are also common, appearing in just over 7% of sentences. The emphasis on these five topics suggests that senior managers devote substantial narrative space to topics emphasizing the benefit of shareholders and firms. As indicated by $TOPIC_{Market}$'s occurrence in roughly 8% of sentences, managers are also careful to make an effort to establish the fairness and comparability of their compensation by

referencing market survey sources. Overall, at the category level, topics related to *accounting*, *aligning interests*, and *pay-for-performance* are more pervasive than other topic categories. At the lower end, the *risk* topic is found in just over 2% of sentences. In identifying which topics are most salient in the CD&A disclosures, the above frequencies provide insights into the narrative themes management believes are most critical for managing shareholder impressions.

5.2 CD&A topics and Changes in Accounting Performance

I now turn to RQ2 and examine how changes in accounting numbers drive the variation in topical emphasis seen in Figure 2.2. To start, Table 2.4 reports summary statistics for all variables included in the sentence-level regressions. Given my use of binary variables to define the topic variables, the mean values for all 19 topic variables are consistent with what I see in Figure 2.2. For example, I see that $TOPIC_{Accting}$, with a mean value of 0.102, is the topic most frequently included in CD&A sentences and that the other 18 topics all have average values ranging from 0.023 to 0.084. For the eight combined topics, $TOPIC_{Combined\text{-}Alignment}$ is the most prominent, with a mean of 0.264. $TOPIC_{Combined\text{-}Risks}$ exhibits the lowest average value among all the combined topics. Concerning the key explanatory variable, *ROA_INCREASE*, I see that the average value is 0.493. For the control variables, I find average values on *ASSETS*, *LEVERAGE,* and *TURNOVER* of 8.069, 0.290, and 1.015, respectively. *MTB* has a mean value of 1.869, the mean value for *PPE* is 0.259, for *CAPEX is* 0.047, and for *EXECPAY* is 1.083. The average *LENGTH* is 4.491.

Table 2.4 Descriptive Statistics for Variables in Regressions

Variables	N	Mean	SD	Min	Max
Accounting					
$TOPIC_{Accting}$	1,980,034	0.102	0.303	0	1
Aligning interests					
$TOPIC_{Incentives}$	1,980,034	0.081	0.273	0	1
$TOPIC_{Voting}$	1,980,034	0.082	0.274	0	1
$TOPIC_{Goals}$	1,980,034	0.084	0.277	0	1
$TOPIC_{Efficiency}$	1,980,034	0.061	0.239	0	1
$TOPIC_{Combined\text{-}Alignment}$	1,980,034	0.264	0.441	0	1
Pay-for-performance					
$TOPIC_{Stock}$	1,980,034	0.073	0.260	0	1
$TOPIC_{Non\text{-}GAAP}$	1,980,034	0.028	0.166	0	1
$TOPIC_{Bonus}$	1,980,034	0.070	0.256	0	1
$TOPIC_{Combined\text{-}Performance}$	1,980,034	0.158	0.365	0	1
Peers					
$TOPIC_{Peer}$	1,980,034	0.055	0.227	0	1
$TOPIC_{Market}$	1,980,034	0.081	0.273	0	1
$TOPIC_{Combined\text{-}Peers}$	1,980,034	0.131	0.337	0	1
Independence					
$TOPIC_{Committee}$	1,980,034	0.058	0.234	0	1
$TOPIC_{Advisor}$	1,980,034	0.062	0.240	0	1
$TOPIC_{Combined\text{-}Independence}$	1,980,034	0.115	0.319	0	1
Basic introduction					
$TOPIC_{Elements}$	1,980,034	0.064	0.245	0	1
$TOPIC_{Introduction}$	1,980,034	0.049	0.216	0	1
$TOPIC_{Overview}$	1,980,034	0.083	0.275	0	1
$TOPIC_{Termination}$	1,980,034	0.072	0.258	0	1
$TOPIC_{Combined\text{-}Intro}$	1,980,034	0.237	0.425	0	1
Risk management					
$TOPIC_{Risk}$	1,980,034	0.023	0.151	0	1
$TOPIC_{Clawback}$	1,980,034	0.056	0.229	0	1
$TOPIC_{Combined\text{-}Risks}$	1,980,034	0.077	0.266	0	1
Compliance					
$TOPIC_{Compliance}$	1,980,034	0.067	0.250	0	1
Control variables					
ROA_INCREASE	1,980,034	0.493	0.500	0	1
ASSETS	1,980,034	8.069	1.713	3.576	12.685
LEVERAGE	1,980,034	0.290	0.666	0	33.008
TURNOVER	1,980,034	1.015	1.454	0.002	99.553
MTB	1,980,034	1.869	3.180	-28.767	179.368
PPE	1,980,034	0.259	0.279	0	4.771
CAPEX	1,980,034	0.047	0.066	0	2.787
EXECPAY	1,980,034	1.083	2.132	0.001	59.674
LENGTH	1,980,034	4.491	0.719	0.729	5.103

Note: Table 2.4 presents summary statistics for all the variables in my sentence-level regressions. The sentence-level sample comprises 1,980,034 sentences over fiscal years 2006 to 2015. The 19 topic adoption variables are predictors and are presented according to eight broader topical categories (shown at the top of each category and denoted by "Combined" in the variable name). I operationalize ROA_INCREASE as the dependent variable. Variable definitions are presented in the Appendix A.

Table 2.5 presents the estimates from the LASSO regressions. As noted earlier, LASSO is a machine-learning-based technique that allows me to explore the relationship between increases in ROA and topic adoption. Essentially, distinct from the causal order of variables in a typical regression, the LASSO seeks to predict the change in ROA that occurred (increase or no increase) based on the topics that managers subsequently chose to emphasize in their CD&A disclosures.[10] On an operational level, LASSO regression allows me to simultaneously test the relationship between *ROA_INCREASE* and 19 topics plus controls. Similar to a typical OLS regression, a positive (negative) coefficient on a topic indicates that the presence of this topic is more (less) predictive of an increase in ROA in the current year; however, different from a typical regression, LASSO's penalization method pushes less powerful coefficients towards zero, producing a more parsimonious model where only "significant" coefficients are shown in the table.[11] Column (1) in Table 2.5 presents results for the 19 individual topic adoption variables, while Column (2) shows results for the eight combined topic variables. I analyze results for both columns by each of the eight combined categories in turn.

[10] Accordingly, in Table 2.5, while ROA_INCREASE superficially appears to be a "dependent" variable, it is actually my explanatory variable as I expect it to impact senior managers' justification of their compensation disclosure narratives. This is in line with other topic modeling research (Morgan-Lopez et al. 2017; Neu & Saxton, 2023; Rao et al. 2010) that use LASSO to predict a precursor variable (in my case, the level of *ROA_INCREASE*) based on the topics that are included in the document. The LASSO also provides other benefits in my context. To start, LASSO assists in identifying the topics most strongly associated with changes in ROA while reducing overfitting in a setting with a large number of candidate predictors. Its penalization approach also helps address multicollinearity among the topic variables, facilitating identification of the strongest systematic associations between ROA changes and CD&A topic emphasis.

[11] As noted earlier, one difference between LASSO regression and regular regressions is that LASSO regression results only include coefficients for variables that are significant in results by shrinking some coefficients to zero. In the main results, however, all the topic variables and controls are significant.

Table 2.5 LASSO-Based Estimates of the Relationship between ROA Increase and Topic Adoption Variables

	Dependent variable = ROA_INCREASE	
	(1)	(2)
	IVs are 19 individual topic variables	*IVs are 8 combined topic variables*
Intercept	0.5387	0.5383
Accounting		
$TOPIC_{Accting}$	-0.0006	-0.0001
Aligning interests		
$TOPIC_{Incentives}$	-0.0005	
$TOPIC_{Voting}$	0.0006	
$TOPIC_{Goals}$	0.0018	
$TOPIC_{Efficiency}$	0.0070	
$TOPIC_{Combined\text{-}Alignment}$		0.0027
Pay-for-performance		
$TOPIC_{Stock}$	-0.0048	
$TOPIC_{Non\text{-}GAAP}$	-0.0042	
$TOPIC_{Bonus}$	0.0049	
$TOPIC_{Combined\text{-}Performance}$		-0.0012
Peers		
$TOPIC_{Peer}$	-0.0035	
$TOPIC_{Market}$	-0.0010	
$TOPIC_{Combined\text{-}Peers}$		-0.0020
Independence		
TOPICCommittee	-0.0057	
TOPICAdvisor	0.0010	
$TOPIC_{Combined\text{-}Independence}$		-0.0027
Basic introduction		
$TOPIC_{Elements}$	-0.0074	
$TOPIC_{Introduction}$	0.0058	
$TOPIC_{Overview}$	0.0030	
$TOPIC_{Termination}$	0.0035	
$TOPIC_{Combined\text{-}Intro}$		0.0020
Risk management		
$TOPIC_{Risk}$	-0.0022	
$TOPIC_{Clawback}$	-0.0041	
$TOPIC_{Combined\text{-}Risks}$		-0.0038
Compliance		
$TOPIC_{Compliance}$	0.0050	0.0045
Control variables		
ASSETS	-0.0041	-0.0040
LEVERAGE	0.0035	0.0035
TURNOVER	0.0093	0.0093
MTB	-0.0015	-0.0015
PPE	0.0240	0.0241
CAPEX	-0.0570	-0.0572
EXECPAY	0.0010	0.0010
LENGTH	-0.0009	-0.0009
YEAR FIXED EFFECTS	Yes	Yes
INDUSTRY FIXED EFFECTS	Yes	Yes
N	1,980,034	1,980,034

Note: Table 2.5 reports the results of sentence-level LASSO regressions of ROA INCREASE on topic adoption variables. The sample comprises 1,980,034 sentences from 13,057 CD&A disclosures over fiscal years 2006 to 2015. Column (1) presents LASSO estimates of 19 individual topic adoption variables while column (2) presents LASSO estimates of eight combined topic adoption variables. All the variables in the table are significant. A positive (negative) coefficient for a topic adoption variable indicates a positive (negative) association between ROA increase and the likelihood of the topic appearing in a CD&A sentence. As suggested by Hastie et al. (2023), I tune the parameter 'lambda' to 'lambda.min' when performing LASSO regressions.

Starting with the accounting category, the findings for both columns (1) and (2) suggest that top management is less likely to use the accounting topic ($TOPIC_{Accting}$) to frame CD&A narratives when there is an increase in ROA. This suggests that managers de-emphasize raw financial terms when performance improves, perhaps assuming such accounting information is already visible to investors elsewhere. This could also suggest that, in periods of strong financial performance, managers believe that stakeholders are more concerned with how the benefits of this performance are shared among executives and shareholders, rather than with the detailed accounting information itself.

Within the *aligning interests* category, while the coefficient in column (2) for the combined variable $TOPIC_{Combined\text{-}Alignment}$ is positive, the findings in column (1) for the individual topics are more nuanced: the coefficients for $TOPIC_{Voting}$, $TOPIC_{Goals}$, and $TOPIC_{Efficiency}$ are positive, but the coefficient on $TOPIC_{Incentives}$ is negative. This indicates that, in the presence of ROA increase, senior managers prefer to talk more about shareholders' voting, firms' ultimate goals, and short-term incentives to mitigate concerns about managerial opportunism. Interestingly, managers reduce emphasis on $TOPIC_{Incentives}$ when ROA increases, possibly because short-term results are more salient to stakeholders during periods of improved performance.

Next, the *pay-for-performance* category is negatively associated with increases in ROA, suggesting that the overall probability of leveraging a pay-for-performance narrative tends to decrease with ROA growth. As seen in column (1), this primarily stems from the lower probability of employing $TOPIC_{Stock}$ and $TOPIC_{Non\text{-}GAAP}$, with only $TOPIC_{Bonus}$ obtaining a positive coefficient.

The reduced emphasis on stock grants and non-GAAP metrics may reflect managers' belief that, in strong-performing years, such technical disclosures are less necessary to justify compensation. Instead, they highlight bonuses, which offer a more direct and intuitive link between performance and reward, making the narrative appear straightforward.[12]

The coefficients for the *peers* topic are negative for the two individual topics, as shown in columns (1), as well as the combined category in column (2). This pattern suggests that senior managers are less likely to rely on $TOPIC_{Peer}$ and $TOPIC_{Market}$ when performance is strong. A plausible reason could be that firms with increasing ROA are inclined to offer higher compensation to executives, which makes managers less inclined to benchmark their remuneration against industry standards.

Next comes the *independence* category. Interestingly, when ROA has increased, firms reduce the discussion of $TOPIC_{Committee}$ while highlighting $TOPIC_{Advisor}$, underscoring the value that management places on external validation of their decisions. Overall, though, the negative coefficient for $TOPIC_{Combined\text{-}Independence}$ in column (2) implies that managers are less prone to discuss the independence of the decision-making process when their firms are performing well, possibly due to a perceived decreased need to justify their decisions in the face of success.

The sixth category of topics, which provides a *basic introduction* about the executive compensation design, is largely positively associated with increases in ROA, with three out of four individual topics in column (1) and the combined category in column (2) obtaining positive coefficients. Emphasizing foundational content may serve to reassure that compensation is structured, principled, and aligned with firm performance. This emphasis becomes especially

[12] This result aligns with Packard et al.'s (2023) findings that firms exhibiting good performance provide fewer performance-related sentences in their CD&A narratives.

salient in strong-performing years, when investors might expect higher compensation levels but also demand assurance that such rewards are fair and based on sound governance practices.

The results for the seventh category of topics, *risk management,* are the opposite: the coefficients on $TOPIC_{Risk}$ and $TOPIC_{Clawback}$ in column (1), as well as on $TOPIC_{Combined\text{-}Risks}$ in column (2), are all negative. This suggests that managers are less likely to view any impression management benefits from discussing risk management in years of ROA increase. Emphasizing risks in a high-performing year could invite doubts about operational sustainability or managerial prudence, counteracting the impression of stability and success that managers seek to project.

Lastly, the positive coefficients for $TOPIC_{Compliance}$ in both columns (1) and (2) indicate that senior managers are more likely to highlight firms' *compliance* with regulations in years of ROA growth. By underscoring adherence to legal standards, they seek to pre-empt concerns about regulatory breaches and reinforce the legitimacy of executive compensation in the context of strong financial gains.

In sum, when ROA increases, top management relies more heavily on *aligning interests*, *basic introduction,* and *compliance* categories to justify their compensation and maintain investor trust. Simultaneously, they de-emphasize *accounting*, *pay-for-performance*, *peers*, *independence*, and *risk-related* categories to deflect scrutiny and reinforce the narrative of managerial success. These patterns suggest that under strong performance, managers shift emphasis toward themes that signal fairness, governance, and regulatory compliance, while downplaying technical details and justifications that might attract scrutiny.

6. Additional Analyses and Robustness Tests

I also undertake a variety of additional analyses to further explore insights suggested by the main analyses as well as sensitivity analyses to verify the robustness of the findings to alternative specifications.

6.1 Miscellaneous Robustness Tests

I conduct three robustness tests to assess the stability of the findings. First, to address potential biases due to missing data in the control variables, I rerun the LASSO regressions using only observations with complete data. The untabulated results align closely with the primary findings. Second, following Farshadfar et al. (2025) and Packard et al. (2023), I include additional control variables that might affect both disclosures and performance; specifically, CEO age, CEO tenure, and CEO duality. The inclusion of these variables did not materially change the results. Lastly, I change the number of topics in the STM model to 15 topics and 20 topics. While some coefficients changed slightly in magnitude or direction, the types of justifications and the substance of topics remained largely unchanged across these different topic models.

6.2 Alternative Measure of Topic

In the main regressions, my measures allow for more than one topic per sentence. To assess the robustness of the findings, I also test an alternative measure that identifies the single most-prevalent topic in each sentence based on which of the topics has the highest FREX word count in the sentence. I apply this approach for both the 19 individual topics and the eight combined categories and then re-run the LASSO regressions. The untabulated results are consistent with the main analyses in terms of the sign and significance of the coefficients.

6.3 Results of Filing-Level LASSO Regressions

To validate my choice of the sentence-level unit of analysis, I also perform regressions with filing-level measures. Specifically, I generate a series of topic proportion variables where, for each topic *i*, $TOPIC_PROP_i$ is the total number of FREX words for the topic in the entire CD&A section scaled by the total word count in the CD&A. Following the main analyses, I also create analogous variables for the eight combined topic variables. Table 2.6 reports summary statistics for all variables in the filing-level regressions. The averages for *ROA_INCREASE* and control variables are similar to those reported in Table 2.4. The mean values of individual topic proportion variables are slightly different from those in Table 2.4 due to the changed unit of analysis and measurement approach.

Table 2.6 Descriptive Statistics for Filing-Level Tests

Variables	N	Mean	SD	Min	Max
Accounting					
$TOPIC_PROP_{Accting}$	13,057	0.005	0.003	0	0.031
Aligning interests					
$TOPIC_PROP_{Incentives}$	13,057	0.003	0.001	0	0.013
$TOPIC_PROP_{Voting}$	13,057	0.004	0.002	0	0.020
$TOPIC_PROP_{Goals}$	13,057	0.004	0.002	0	0.018
$TOPIC_PROP_{Efficiency}$	13,057	0.003	0.002	0	0.015
$TOPIC_PROP_{Combined\text{-}Alignment}$	13,057	0.013	0.004	0	0.033
Pay-for-performance					
$TOPIC_PROP_{Stock}$	13,057	0.003	0.002	0	0.023
$TOPIC_PROP_{Non\text{-}GAAP}$	13,057	0.001	0.001	0	0.012
$TOPIC_PROP_{Bonus}$	13,057	0.003	0.003	0	0.020
$TOPIC_PROP_{Combined\text{-}Performance}$	13,057	0.008	0.004	0	0.027
Peers					
$TOPIC_PROP_{Peer}$	13,057	0.003	0.003	0	0.038
$TOPIC_PROP_{Market}$	13,057	0.004	0.002	0	0.027
$TOPIC_PROP_{Combined\text{-}Peers}$	13,057	0.007	0.004	0	0.045
Independence					
$TOPIC_PROP_{Committee}$	13,057	0.002	0.001	0	0.017
$TOPIC_PROP_{Advisor}$	13,057	0.003	0.002	0	0.018
$TOPIC_PROP_{Combined\text{-}Independence}$	13,057	0.005	0.002	0	0.020
Basic introduction					
$TOPIC_PROP_{Elements}$	13,057	0.003	0.001	0	0.011
$TOPIC_PROP_{Introduction}$	13,057	0.003	0.002	0	0.020
$TOPIC_PROP_{Overview}$	13,057	0.004	0.002	0	0.023
$TOPIC_PROP_{Termination}$	13,057	0.004	0.003	0	0.025

$TOPIC_PROP_{Combined\text{-}Intro}$	13,057	0.013	0.004	0	0.034
Risk management					
$TOPIC_PROP_{Risk}$	13,057	0.001	0.001	0	0.014
$TOPIC_PROP_{Clawback}$	13,057	0.002	0.001	0	0.011
$TOPIC_PROP_{Combined\text{-}Risks}$	13,057	0.003	0.001	0	0.016
Compliance					
$TOPIC_PROP_{Compliance}$	13,057	0.004	0.002	0	0.019
Control variables					
ROA_INCREASE	13,057	0.493	0.500	0	1
ASSETS	13,057	7.944	1.737	3.576	12.685
LEVERAGE	13,057	0.289	0.664	0	33.008
TURNOVER	13,057	1.019	1.399	0.002	99.553
MTB	13,057	1.898	3.227	-28.767	179.368
PPE	13,057	0.260	0.289	0	4.771
CAPEX	13,057	0.047	0.072	0	2.787
EXECPAY	13,057	1.185	2.489	0.001	59.674
LENGTH	13,057	4.295	0.927	0.729	5.103

Note: Table 2.6 presents summary statistics for all the variables in my aggregate-level regressions for 13,057 CD&A documents over fiscal years 2006 to 2015. The topic proportion variables are my main predictors and are presented according to eight broader topical categories.

Table 2.7 presents the results of LASSO regressions on the topic proportion variables. A positive (negative) coefficient for these variables indicates that an increase in ROA is associated with a higher (lower) disclosure level of the topic in the entire CD&A section. The outcomes for the 19 individual topic proportion variables are presented in Column (1), whereas Column (2) shows the results for the eight combined variables. I find fewer significant individual topics in the filing-level regression compared to the sentence-level results, but the direction of the associations remains largely consistent. At the combined level, most categories exhibit similar signs and significance as in the main analyses, lending support to the primary findings. Notably, however, two exceptions emerge: the coefficient for $TOPIC_PROP_{Combined\text{-}Performance}$ turns positive, and the coefficient on $TOPIC_PROP_{Accting}$ loses statistical significance.

Table 2.7 LASSO-Based Estimates of the Relationship between ROA Increase and Topic Proportion Variables

	Dependent variable = ROA_INCREASE	
	(1)	(2)
	IVs are 19 individual topic variables	*IVs are 8 combined topic variables*
Intercept	0.5145	0.4832
Accounting		
$TOPIC_PROP_{Accting}$	-	-
Aligning interests		
$TOPIC_PROP_{Incentives}$	-	
$TOPIC_PROP_{Voting}$	1.3990	
$TOPIC_PROP_{Goals}$	-	
$TOPIC_PROP_{Efficiency}$	5.0204	
$TOPIC_PROP_{Combined\text{-}Alignment}$		2.0476
Pay-for-performance		
$TOPIC_PROP_{Stock}$	-0.8470	
$TOPIC_PROP_{Non\text{-}GAAP}$	-	
$TOPIC_PROP_{Bonus}$	2.4136	
$TOPIC_PROP_{Combined\text{-}Performance}$		0.1616
Peers		
$TOPIC_PROP_{Peer}$	-0.4528	
$TOPIC_PROP_{Market}$	-0.0956	
$TOPIC_PROP_{Combined\text{-}Peers}$		-0.5228
Independence		
$TOPIC_PROP_{Committee}$	-4.1063	
$TOPIC_PROP_{Advisor}$	-	
$TOPIC_PROP_{Combined\text{-}Independence}$		-0.9289
Basic introduction		
$TOPIC_PROP_{Elements}$	-7.3806	
$TOPIC_PROP_{Introduction}$	1.2470	
$TOPIC_PROP_{Overview}$	2.9695	
$TOPIC_PROP_{Termination}$	1.1844	
$TOPIC_PROP_{Combined\text{-}Intro}$		0.7206
Risk management		
$TOPIC_PROP_{Risk}$	-	
$TOPIC_PROP_{Clawback}$	-4.8780	
$TOPIC_PROP_{Combined\text{-}Risks}$		-3.7774
Compliance		
$TOPIC_PROP_{Compliance}$	3.8947	3.9755
Control variables		
ASSETS	-0.0054	-0.0033
LEVERAGE	-	-
TURNOVER	0.0077	0.0078
MTB	-0.0008	-0.0009
PPE	-	-
CAPEX	-	-
EXECPAY	-	-
LENGTH	-	-
YEAR FIXED EFFECTS	Yes	Yes
INDUSTRY FIXED EFFECTS	Yes	Yes
N	13,057	13,057

Note: Table 2.7 reports the results of filing-level LASSO regressions of ROA_INCREASE on topic proportion variables. The sample comprises 13,057 CD&A disclosures over fiscal years 2006 to 2015. Column (1) presents LASSO estimates of 19 individual topic proportion variables while column (2) presents LASSO estimates of eight combined topic proportion variables. Insignificant variables are marked with a "-" in cells. A positive (negative) coefficient for a topic proportion variable indicates a positive (negative) association between ROA increase and the disclosure level of the topic in a CD&A document.

To further check on the robustness of my findings, I follow Beattie's (2014) insights that managers may use "abnormal" language choices as part of impression management strategies. I therefore construct an alternative, "abnormal," measure of topic disclosure. For each topic, I first define an "average FREX word count" variable as the average value of a firm's overall FREX word count of a topic across all prior sample years. The "abnormal" topic disclosure variable $TOPIC_ABN_i$ equals one if the overall FREX word count of topic *i* in a given CD&A filing exceeds its average FREX word count in that firm's prior filings by 20%, and zero otherwise. I then do the same for each of the eight combined topic categories. Table 2.8 shows the results of the filing-level LASSO regressions of *ROA_INCREASE* on these alternative topic disclosure variables. The results for the *aligning interests, independence,* and *compliance* categories are consistent with those in Table 2.5. However, five categories—*accounting, pay-for-performance, peers, basic introduction,* and *risk management*—now show no relationship with ROA increase. These findings indicate that periods of improved performance are associated with greater-than-historical emphasis on selected themes, consistent with the strategic use of narrative topics for impression management.

Table 2.8 LASSO-Based Estimates of the Relationship between ROA Increase and Abnormal Topic Disclosure Variables

	Dependent variable = ROA_INCREASE	
	(1) *IVs are 19 individual topic variables*	(2) *IVs are 8 combined topic variables*
Intercept	0.5176	0.4978
Accounting		
$TOPIC_ABN_{Accting}$	-	-
Aligning interests		
$TOPIC_ABN_{Incentives}$	-0.0052	
$TOPIC_ABN_{Voting}$	0.0057	
$TOPIC_ABN_{Goals}$	0.0029	
$TOPIC_ABN_{Efficiency}$	0.0081	
$TOPIC_ABN_{Combined\text{-}Alignment}$		0.0016
Pay-for-performance		
$TOPIC_ABN_{Stock}$	-0.0156	
$TOPIC_ABN_{Non\text{-}GAAP}$	-	
$TOPIC_ABN_{Bonus}$	0.0130	
$TOPIC_ABN_{Combined\text{-}Performance}$		-
Peers		
$TOPIC_ABN_{Peer}$	-	
$TOPIC_ABN_{Market}$	-	
$TOPIC_ABN_{Combined\text{-}Peers}$		-
Independence		
$TOPIC_ABN_{Committee}$	-0.0144	
$TOPIC_ABN_{Advisor}$	-0.0005	
$TOPIC_ABN_{Combined\text{-}Independence}$		-0.0052
Basic introduction		
$TOPIC_ABN_{Elements}$	-0.0145	
$TOPIC_ABN_{Introduction}$	0.0094	
$TOPIC_ABN_{Overview}$	0.0123	
$TOPIC_ABN_{Termination}$	0.0152	
$TOPIC_ABN_{Combined\text{-}Intro}$	-	-
Risk management		
$TOPIC_ABN_{Risk}$	-0.0007	
$TOPIC_ABN_{Clawback}$	-0.0025	
$TOPIC_ABN_{Combined\text{-}Risks}$		-
Compliance		
$TOPIC_ABN_{Compliance}$	0.0104	0.0077
Control variables		
ASSETS	-0.0047	-0.0017
LEVERAGE	-	-
TURNOVER	0.0078	0.0065
MTB	-0.0010	-0.0003
PPE	-	-
CAPEX	-	-
EXECPAY	-	-
LENGTH	-	-
YEAR FIXED EFFECTS	Yes	Yes
INDUSTRY FIXED EFFECTS	Yes	Yes

N	13,057	13,057

Note: Table 2.8 reports the results of filing-level LASSO regressions of ROA_INCREASE on "abnormal" topic disclosure variables. The sample comprises 13,057 CD&A disclosures over fiscal years 2006 to 2015. Column (1) presents LASSO estimates of 19 individual abnormal topic disclosure variables while column (2) presents LASSO estimates of eight combined abnormal topic disclosure variables. Insignificant variables are marked with a "-" in cells. A positive (negative) coefficient for an abnormal topic disclosure variable indicates a positive (negative) association between ROA increase and the likelihood of above-historical topic emphasis in a CD&A document.

6.4 Alternative Measures of Accounting Number Change

To complement the main tests, I also run regressions using the *decrease* in ROA as the proxy for accounting number change. I assign a value of one on *ROA_DECREASE* when there is a 2.5% or greater decline in ROA compared to the previous year, and zero otherwise. This 2.5% threshold reflects a balanced cutoff, as approximately 45% of firm-year observations in the sample exceed this level of decline, making it neither overly stringent nor overly lenient. The untabulated results show that the signs of the coefficients, for both individual and combined topic variables, are the opposite of those reported in Table 2.5. These findings reinforce the main conclusion that changes in accounting performance are systematically associated with variation in CD&A topic emphasis.

6.5 Relationship between CEO Pay Raises and Topic Choice

Following prior research exploring the link between CEO pay and CD&A readability (Hemmings et al., 2020; Hooghiemstra et al., 2017; Laksmana et al., 2012), I examine the relationship between changes in CEO pay and compensation narratives. Specifically, I re-run the LASSO regressions from Table 2.5 replacing *ROA_INCREASE with CEOPAY_INCREASE,* a binary variable equal to one if total CEO compensation increases in the current year, zero otherwise. Table 2.9 presents the results for both individual and combined topics. Save for three individual topics – $TOPIC_{Incentives}$, $TOPIC_{Introduction}$, and $TOPIC_{Clawback}$ – and the combined topic $TOPIC_{Combined\text{-}Risks}$, topic selection is related to the increase in CEO pay. Overall, increases in CEO

compensation are associated with greater emphasis on topics in the combined categories *accounting*, *aligning interests, pay-for-performance, peers,* and *compliance* and lower emphasis on the topics in the *independence* and *basic introduction* categories.

Table 2.9 LASSO-Based Estimates of the Relationship between CEO Pay Increase and Topic Adoption Variables

	Dependent variable = CEOPAY_INCREASE	
	(1) *IVs are 19 individual topic variables*	(2) *IVs are 8 combined topic variables*
Intercept	0.6287	0.6289
Accounting		
$TOPIC_{Accting}$	0.0004	0.00001
Aligning interests		
$TOPIC_{Incentives}$	-	
$TOPIC_{Voting}$	0.0018	
$TOPIC_{Goals}$	0.0009	
$TOPIC_{Efficiency}$	-0.0041	
$TOPIC_{Combined\text{-}Alignment}$		0.0008
Pay-for-performance		
$TOPIC_{Stock}$	0.0039	
$TOPIC_{Non\text{-}GAAP}$	0.0008	
$TOPIC_{Bonus}$	-0.0008	
$TOPIC_{Combined\text{-}Performance}$		0.0016
Peers		
$TOPIC_{Peer}$	-0.0033	
$TOPIC_{Market}$	0.0041	
$TOPIC_{Combined\text{-}Peers}$		0.0010
Independence		
TOPICCommittee	-0.0015	
TOPICAdvisor	-0.0001	
$TOPIC_{Combined\text{-}Independence}$		-0.0007
Basic introduction		
$TOPIC_{Elements}$	-0.0010	
$TOPIC_{Introduction}$	-	
$TOPIC_{Overview}$	0.0009	
$TOPIC_{Termination}$	-0.0033	
$TOPIC_{Combined\text{-}Intro}$		-0.0011
Risk management		
$TOPIC_{Risk}$	-0.00002	
$TOPIC_{Clawback}$	-	
$TOPIC_{Combined\text{-}Risks}$		-
Compliance		
$TOPIC_{Compliance}$	0.0014	0.0014
Control variables		
ASSETS	-0.0081	-0.0081
LEVERAGE	0.0134	0.0134
TURNOVER	0.0090	0.0090
MTB	-0.0033	-0.0033
PPE	0.0568	0.0570

CAPEX	0.0937	0.0941
LENGTH	0.0032	0.0032
YEAR FIXED EFFECTS	Yes	Yes
INDUSTRY FIXED EFFECTS	Yes	Yes
N	1,883,755	1,883,755

Note: Table 2.9 reports the results of sentence-level LASSO regressions of CEOPAY INCREASE on topic adoption variables. The sample comprises 1,883,755 sentences from 12,287 CD&A disclosures over fiscal years 2006 to 2015. The sample is slightly lower than the main regression in Table 2.5 owing to firms missing Execucomp data. Column (1) presents LASSO estimates of 19 individual topic adoption variables while column (2) presents LASSO estimates of eight combined topic adoption variables. Insignificant topics are marked with a "-" in cells. A positive (negative) coefficient for a topic adoption variable indicates a positive (negative) association between CEO compensation increases and the likelihood of topic adoption.

In untabulated tests I also examine whether topic emphasis varies with the magnitude of CEO pay increases. I create two binary variables indicating CEO pay increases in the highest quartile and the lowest quartile, respectively. The results show a narrower range of emphasized CD&A topics when pay increases are more modest, compared with the broader range associated with larger compensation increases. The varying signs and magnitudes of coefficients across the compensation tiers suggest that the extent of CEO pay raises is associated with differences in topic emphasis.

6.6 Earnings Management and CD&A Topic Adoption

While the main analyses examine topic variation in relation to changes in accounting performance, they do not explicitly consider firms whose reported earnings just meet or slightly exceed analysts' forecasts. To address this issue, I examine the association between earnings management, proxied by whether a firm just meets or beats analysts' consensus forecasts, and topic adoption. I assign the variable *JMB* a value of one if a firm's earnings just meet or slightly exceed analysts' consensus earnings forecast by 2 cents or less, and zero otherwise. After dropping observations with missing values, the final sample consists of 262,933 sentences. The results in Table 2.10 show that all combined topics, except for $TOPIC_{Accting}$ and $TOPIC_{Combined\text{-}Risks}$, are negatively related to earnings management. This suggests that when managers engage in earnings

management, they are less likely to justify compensation design by employing otherwise common narrative themes such as those in the *aligning interests, pay-for-performance, peers*, and *independence* categories. Overall, earnings management is associated with reduced emphasis on several common compensation-justification themes in CD&A disclosures.

Table 2.10 LASSO-Based Estimates of the Relationship between Earnings Management and Topic Adoption Variables

	Dependent variable = JMB	
	(1) *IVs are 19 individual topic variables*	(2) *IVs are 8 combined topic variables*
Intercept	0.1789	0.1797
Accounting		
$TOPIC_{Accting}$	0.0063	0.0063
Aligning interests		
$TOPIC_{Incentives}$	-0.0072	
$TOPIC_{Voting}$	-0.0023	
$TOPIC_{Goals}$	0.0041	
$TOPIC_{Efficiency}$	-0.0055	
$TOPIC_{Combined\text{-}Alignment}$		-0.0036
Pay-for-performance		
$TOPIC_{Stock}$	0.0009	
$TOPIC_{Non\text{-}GAAP}$	0.0026	
$TOPIC_{Bonus}$	-0.0091	
$TOPIC_{Combined\text{-}Performance}$		-0.0048
Peers		
$TOPIC_{Peer}$	-0.0190	
$TOPIC_{Market}$	-0.0159	
$TOPIC_{Combined\text{-}Peers}$		-0.0180
Independence		
TOPICCommittee	-0.0103	
TOPICAdvisor	-0.0079	
$TOPIC_{Combined\text{-}Independence}$		-0.0106
Basic introduction		
$TOPIC_{Elements}$	-0.0068	
$TOPIC_{Introduction}$	-0.0093	
$TOPIC_{Overview}$	0.0002	
$TOPIC_{Termination}$	0.0121	
$TOPIC_{Combined\text{-}Intro}$		-0.0006
Risk management		
$TOPIC_{Risk}$	-0.0018	
$TOPIC_{Clawback}$	0.0062	
$TOPIC_{Combined\text{-}Risks}$		0.0039
Compliance		
$TOPIC_{Compliance}$	-0.0015	-0.0014
Control variables		
ASSETS	-0.0020	-0.0021
LEVERAGE	-0.0008	-0.0007
TURNOVER	-0.0092	-0.0092

MTB	0.0048	0.0048
PPE	-0.1225	-0.1224
CAPEX	0.1154	0.1159
EXECPAY	-0.0033	-0.0033
LENGTH	-0.0181	-0.0182
YEAR FIXED EFFECTS	YES	YES
INDUSTRY FIXED EFFECTS	YES	YES
N	262,933	262,933

Note: Table 2.10 reports the results of sentence-level LASSO regressions of JMB (an indicator variable assigned a value of one if a firm's earnings just meet or slightly exceed analysts' consensus earnings forecast by 2 cents or less, and zero otherwise) on topic adoption variables. The sample comprises 262,933 sentences from 1,718 CD&A disclosures over fiscal years 2006 to 2015. Column (1) presents LASSO estimates of 19 individual topic adoption variables while column (2) presents LASSO estimates of eight combined topic adoption variables. All the variables shown in the table are significant. A positive (negative) coefficient for a topic adoption variable indicates a positive (negative) association between the earnings-management proxy and the likelihood of topic adoption.

6.7 CD&A Topic Choices and Market Reactions

The evidence of systematic variation in CD&A topic choices raises a related question: whether these topic choices are associated with market reactions around the filing date. To examine this relationship, I use the filing-level topic proportion variables from Section 5.3 and analyze firms' three-day cumulative abnormal returns (*CAR*) around their CD&A filing dates. The results in Table 2.11 show that greater emphasis on several CD&A topics related to clarity, formality, and regulatory adherence, including compensation package components, market survey sources, CD&A overview, and compliance, is associated with more positive abnormal returns around the filing date. In contrast, greater emphasis on topics related to internal incentive mechanisms, such as long-term incentives and shareholder voting, is associated with more negative abnormal returns. Overall, these findings indicate that the association between CD&A topic emphasis and abnormal returns varies across the substantive content of the disclosure.

Table 2.11 LASSO-Based Estimates of the Relationship between Abnormal Returns and Topic Proportion Variables

	Dependent variable = CAR	
	(1)	(2)
	IVs are 19 individual topic variables	*IVs are 8 combined topic variables*
Intercept	0.0013	0.0017
Accounting		
$TOPIC_PROP_{Accting}$	-	-0.0444
Aligning interests		
$TOPIC_PROP_{Incentives}$	-0.1769	
$TOPIC_PROP_{Voting}$	-0.0106	
$TOPIC_PROP_{Goals}$	-	
$TOPIC_PROP_{Efficiency}$	-	
$TOPIC_PROP_{Combined\text{-}Alignment}$		-0.0091
Pay-for-performance		
$TOPIC_PROP_{Stock}$	0.0207	
$TOPIC_PROP_{Non\text{-}GAAP}$	-	
$TOPIC_PROP_{Bonus}$	-0.1663	
$TOPIC_PROP_{Combined\text{-}Performance}$		-0.0109
Peers		
$TOPIC_PROP_{Peer}$	-	
$TOPIC_PROP_{Market}$	0.0740	
$TOPIC_PROP_{Combined\text{-}Peers}$		-
Independence		
$TOPIC_PROP_{Committee}$	-	
$TOPIC_PROP_{Advisor}$	-	
$TOPIC_PROP_{Combined\text{-}Independence}$		-
Basic introduction		
$TOPIC_PROP_{Elements}$	0.0688	
$TOPIC_PROP_{Introduction}$	-	
$TOPIC_PROP_{Overview}$	0.1668	
$TOPIC_PROP_{Termination}$	-	
$TOPIC_PROP_{Combined\text{-}Intro}$		0.0389
Risk management		
$TOPIC_PROP_{Risk}$	-	
$TOPIC_PROP_{Clawback}$	-0.0085	
$TOPIC_PROP_{Combined\text{-}Risks}$		-0.0177
Compliance		
$TOPIC_PROP_{Compliance}$	0.0318	0.0678
Control variables		
ASSETS	-	-0.0001
LEVERAGE	0.0006	0.0009
TURNOVER	-	-0.0001
MTB	-0.0001	-0.0002
PPE	0.0007	0.0015
CAPEX	-	-
EXECPAY	-	-0.00002
LENGTH	-	-
YEAR FIXED EFFECTS	Yes	Yes
INDUSTRY FIXED EFFECTS	Yes	Yes
N	10,918	10,918

Note: Table 2.11 reports the results of filing-level LASSO regressions of CAR (three-day cumulative abnormal returns) on topic proportion variables. The sample comprises 10,918 CD&A disclosures over fiscal years 2006 to 2015, reflecting the exclusion of firm-year observations with missing return data. Column (1) presents LASSO estimates of 19 individual topic proportion variables while column (2) presents estimates of eight combined topic proportion variables. Insignificant variables are marked with a "-" in cells. A positive (negative) coefficient for a topic proportion variable indicates that greater topic emphasis is associated with more positive (negative) abnormal returns around the CD&A filing date.

6.8 Topics and Future Accounting Number Change

Lastly, I examine whether current-year CD&A topic choices are associated with changes in future accounting performance. Prior research has documented associations between textual characteristics of corporate disclosures, such as tone and readability, and future financial performance (Huang et al., 2014; Li, 2008). Building on this literature, I examine whether current-year CD&A topics contain information associated with subsequent changes in accounting numbers. Table 2.12 presents the results of LASSO regressions relating current-year topic adoption variables to next year's ROA increase. Overall, I find that categories *aligning interests, pay for performance, peers,* and *risk management* are all positively related to future ROA increase. These findings indicate that current-year CD&A topic choices are systematically associated with subsequent financial performance.

Table 2.12 LASSO-Based Estimates of the Relationship between Topic Adoption Variables and Future Accounting Number Change

	Dependent variable = ROA_INCREASE_LEAD	
	(1)	(2)
	IVs are 19 individual topic variables	*IVs are 8 combined topic variables*
Intercept	0.5787	0.5781
Accounting		
$TOPIC_{Accting}$	-0.0017	-0.0010
Aligning interests		
$TOPIC_{Incentives}$	0.0004	
$TOPIC_{Voting}$	-0.0001	
$TOPIC_{Goals}$	-0.0005	
$TOPIC_{Efficiency}$	0.0027	
$TOPIC_{Combined\text{-}Alignment}$		0.0006
Pay-for-performance		

$TOPIC_{Stock}$	-0.0023	
$TOPIC_{Non\text{-}GAAP}$	0.0027	
$TOPIC_{Bonus}$	0.0047	
$TOPIC_{Combined\text{-}Performance}$		0.0015
Peers		
$TOPIC_{Peer}$	0.0004	
$TOPIC_{Market}$	0.0012	
$TOPIC_{Combined\text{-}Peers}$		0.0008
Independence		
TOPICCommittee	-0.0017	
TOPICAdvisor	-0.0027	
$TOPIC_{Combined\text{-}Independence}$		-0.0024
Basic introduction		
$TOPIC_{Elements}$	-0.0010	
$TOPIC_{Introduction}$	-	
$TOPIC_{Overview}$	-0.0015	
$TOPIC_{Termination}$	-0.0003	
$TOPIC_{Combined\text{-}Intro}$		-0.0010
Risk management		
$TOPIC_{Risk}$	0.0033	
$TOPIC_{Clawback}$	0.0021	
$TOPIC_{Combined\text{-}Risks}$		0.0029
Compliance		
$TOPIC_{Compliance}$	-0.0006	-0.0006
Control variables		
ASSETS	-0.0022	-0.0022
LEVERAGE	0.0192	0.0191
TURNOVER	-0.0072	-0.0072
MTB	-0.0058	-0.0058
PPE	-0.0146	-0.0144
CAPEX	-0.5780	-0.5776
EXECPAY	0.0045	0.0044
LENGTH	-0.0067	-0.0067
YEAR FIXED EFFECTS	Yes	Yes
INDUSTRY FIXED EFFECTS	Yes	Yes
N	1,980,034	1,980,034

Note: Table 2.12 reports the results of sentence-level LASSO regressions of ROA INCREASE LEAD on topic adoption variables. The sample comprises 1,980,034 sentences from 13,057 CD&A disclosures over fiscal years 2006 to 2015. Column (1) presents LASSO estimates of 19 individual topic adoption variables while column (2) presents LASSO estimates of eight combined topic adoption variables. Insignificant variables are marked with a "-" in cells. A positive (negative) coefficient for a topic adoption variable indicates a positive (negative) association between topic adoption and ROA increase in the following year.

7. Discussion

Drawing on the lens of impression management (Aerts, 2005; Godfrey et al., 2003), this study examines the topical content of CD&A disclosures and their relationship with changes in

accounting numbers using 1,980,034 sentences from 13,057 CD&A disclosures. Extending prior research on CD&A disclosures focusing on basic textual characteristics such as tone (Balsam et al., 2016) and readability (Laksmana et al., 2012), this study delves into the thematic composition of these disclosures. The first research question seeks to determine the topics that management emphasizes in their CD&A disclosures. Using an LDA-based topic modeling approach, I identify 19 conceptually significant topics organized into eight broad categories: *accounting, aligning interests, pay-for-performance, peers, independence, basic introduction, risk management*, and *compliance*. Of these, the topics in the *accounting, aligning interests, pay-for-performance* categories are more pervasive than others. The answers to RQ1 support and extend prior research in such areas as the purpose of payment incentives (e.g., Kuang & Qin, 2009), pay-performance sensitivity (e.g., Gipper, 2021), firms' emphasis on peer companies (Cadman et al., 2021), compensation committees (Laksmana, 2008), and clawback enforcement (Cha et al., 2023).

Turning to the second research question I examine whether CD&A topic choices vary systematically with changes in key accounting numbers. The LASSO regressions show that an increase in ROA is associated with a greater emphasis on most topics in the *aligning interests* and *compliance* categories. These findings are consistent with the notion of impression management, as improved performance is associated with greater emphasis on selected CD&A themes related to executive compensation. From an impression management perspective, such narratives provide an important means through which firms frame the fairness and reasonableness of executive pay for investors.

My study has certain limitations. To start, while the sample contains 10 years of data, it ends in 2015 and may not fully capture more recent shifts in compensation practices or disclosure norms. Future research could extend the timeline and explore how topical choices change over

time. Second, while I take steps to reduce endogeneity concerns, I cannot completely rule out omitted variable bias in the relationship between performance changes and topic use. Accordingly, the relationships documented in this study should be interpreted as associations rather than causal effects. Future research could adopt quasi-experimental designs, instrumental variable strategies, or other approaches to examine the directionality and causality of these relationships. Lastly, as in all LDA topic modeling approaches, it is up to the researchers to manually create the labels for each topic and classify them into broader categories. The chosen topic labels could thus reflect unconscious researcher biases. Future studies might employ more advanced machine learning techniques to refine the topical structure.

The study offers important implications for investors and regulators. For investors, understanding the thematic structure of CD&A disclosures can improve their ability to assess whether executive compensation is meaningfully aligned with firm performance. For regulators, the findings underscore the importance of evaluating not only the completeness of disclosures but also their substantive content and emphasis. As narrative disclosures provide an important context in which executive compensation is represented, greater attention to the depth and focus of these narratives may provide additional insight into transparency and accountability in executive pay.

Overall, this study shows that CD&A disclosures contain systematic variation in substantive topic emphasis that is associated with changes in accounting performance. By linking topic choices to changing accounting performance, the study highlights substantive topic selection as an additional dimension of narrative disclosure relevant to impression management research. These insights deepen the understanding of how narrative framing complements financial information reported by firms. As narratives become an increasingly powerful medium for firms, understanding their structure is essential for decoding the stories firms choose to tell.

Chapter 3: The Dynamics of Voluntary Social Disclosures in Management Information Circulars

Abstract

This study develops a sentence-level lifecycle framework to examine how voluntary social disclosures evolve over time in Canadian management information circulars (MICs). Drawing on 80,890 voluntary social sentences from 12,774 filings across 2,408 firms spanning 1997 to 2023, I apply FinBERT-ESG embeddings and BERTopic to identify five dominant social topics and classify disclosure content as introduced, carried forward, or abandoned across adjacent firm-years. The findings show that lifecycle stages differ systematically in their substantive content: introduced disclosure tends to reflect new developments and organizational responses, carried-forward disclosure centers more on established systems and commitments, and abandoned disclosure is more often episodic or time-bound. Filing-level analyses further show that disclosure change varies with topic composition, with greater emphasis on Labour Relations and Workplace Health and Safety associated with higher overall turnover. Greater institutional alignment is associated with greater disclosure continuity overall, although this relationship varies with the relative prominence of selected social topics. The study contributes a lifecycle perspective that extends social disclosure research from what firms make visible to how that visibility is sustained and reconfigured over time.

Keywords

Social disclosures; disclosure dynamics; institutional expectations; management information circulars; textual analysis

1. Introduction

The prevailing corporate emphasis on market-driven metrics of value creation and profitability has fostered an economic paradigm that prioritizes growth in profits over broader forms of organizational development (van Bommel, 2014). Yet prior accounting research cautions that a narrow focus on profit maximization can result in substantial social harms ranging from poor working conditions, pollution, and humanitarian disasters (Safari and Parker, 2024). These concerns have prompted growing stakeholder demands for a high "quality of growth"—development that is robust, equitable, and sustainable while generating tangible social benefits such as improved living standards and poverty reduction (Lowe et al., 2020, p. 762). These shifting expectations have expanded the scope of corporate responsibility, pressuring firms to publicly account for their social impacts. Social disclosure has thus emerged as a key channel for signaling organizational accountability (Cho et al., 2015; Patten, 1991).

The growing importance of social disclosure has prompted a substantial body of research examining how social reporting evolves over time. Early work focused on broad shifts in disclosure volume, documenting changes in the amount of social information disclosed in response to heightened stakeholder attention and legitimacy pressures (Gray et al., 1995; Tilling and Tilt, 2010). More recent studies have expanded this focus to examine other dimensions of disclosure change over time, including specificity, rhetorical framing, and narrative style (Bellucci et al., 2021; Lin et al., 2024; Soobaroyen and Mahadeo, 2016). For example, Lin et al. (2024) show that social disclosures have become longer but also more formulaic and less specific over time, suggesting increasing standardization in social reporting practices. Bellucci et al. (2021), meanwhile, demonstrate how firms revise their social messaging following scandals, highlighting the role of disclosure in reputation repair and legitimacy management. Collectively, these studies show that

social disclosure is not static but continually adapts to changing organizational circumstances and stakeholder expectations.

While this literature provides valuable insights into disclosure change, it primarily examines change through aggregate shifts in disclosure characteristics or responses to discrete events. As a result, it offers limited insight into whether and how specific disclosure content is reproduced across successive reporting periods. This limitation is important because disclosure continuity/discontinuity may reveal different aspects of corporate communication than aggregate changes in disclosure volume, tone, or specificity alone. Understanding how firms introduce, carry forward, and abandon social narratives therefore remains an important but underexplored aspect of voluntary social disclosure.

This study addresses this gap by examining year-over-year changes in voluntary social sentences within annual management information circulars (MICs), mandatory documents that Canadian publicly listed firms must distribute to shareholders before annual general meetings.[13] Required under *National Instrument (NI) 51-102*, MICs provide shareholders with the necessary information to make informed decisions on voting matters such as board elections, executive compensation, auditor appointments, and shareholder proposals.[14]

While MICs are designed to report governance information, they also include some mandated social disclosures, specifically regarding gender diversity information on the board and

[13] The Canadian context offers several advantages for my study. First, the Canadian regulatory framework has remained relatively stable across the study period (1997-2023), with minimal changes to social disclosure requirements (primarily the addition of gender diversity disclosures in 2014). This stability enables us to track disclosure evolution without major regulatory disruptions that might confound my analysis of voluntary choices. Second, Canada's resource-based economy and active civil society expose firms to heightened social scrutiny (Thorne *et al.,* 2014). This creates pressure for firms to disclose social information voluntarily. Finally, the Canadian market features robust corporate governance practices compared to other developed markets (Doidge *et al.,* 2007), enhancing the relevance of the setting for examining voluntary social disclosure.

[14] For further information on NI 51-102, see Ontario Securities Commission's website: Continuous disclosure obligations", available at: https://www.osc.ca/en/securities-law/instruments-rules-policies/5/51-102.

in executive officer positions. Beyond these required elements, there are no legal obligations for firms to report on other social topics in their MICs. In other words, reporting on non-discrimination, training and education, labor relations, occupational health and safety, local communities, customer health and safety, public policy, and other social topics in the MICs is wholly voluntary (Global Reporting Initiative Standards, 2016). Despite this minimal regulatory guidance, prior research has documented that firms frequently do choose to include additional social information in MICs, a practice that has been interpreted as consistent with legitimacy-driven responses to stakeholder expectations (Baier et al., 2020; Baker et al., 2024). This distinctive context, characterized by limited mandatory requirements with firms' adoption of voluntary additions, makes MICs a valuable setting for exploring how voluntary social content evolves over time within mandatory filings.

To examine voluntary social disclosure in MICs, this study draws on three theoretical constructs: visibility, temporality, and institutional expectations. Visibility directs attention to which social issues firms choose to foreground within their disclosures (Burchell et al., 1980; Hopwood, 1987; Preston et al., 1996). Temporality examines how that visibility unfolds over time. It examines how certain social messages are continuously reproduced across reporting cycles while others are introduced and subsequently abandoned (Ezzamel and Robson, 1995; Parker, 2004). Institutional expectations examines how alignment with broader social reporting frameworks is associated with the temporal dynamics of disclosure (DiMaggio and Powell, 1983). Together, these constructs motivate three research questions. First, RQ1: *What are the dominant voluntary social topics in Canadian MICs?* Second, RQ2: *How do disclosure lifecycle patterns vary across dominant voluntary social topics?* Third, RQ3: *What is the relationship between institutional alignment and disclosure lifecycle outcomes?*

The study draws on MIC filings obtained from Avantis AI spanning 1997 to 2023. I first identify voluntary social sentences using a dictionary-based procedure, apply FinBERT-ESG embeddings and BERTopic to identify dominant social topics, and classify sentence-level disclosure content as introduced, carried forward, or abandoned by comparing semantic similarity across adjacent firm-years. I then use filing-level post-LASSO OLS models to examine how topic composition and institutional alignment are associated with disclosure lifecycle outcomes.

The findings show that voluntary social disclosure in MICs is organized around five dominant topics: Community and Philanthropy, Workplace Health and Safety, DEI Initiatives, Indigenous Community Engagement, and Labour Relations. The lifecycle analysis reveals systematic differences in the substantive content associated with different lifecycle stages. Newly introduced disclosure tends to concern new developments and organizational responses, carried-forward disclosure tends to center on established systems and commitments, and abandoned disclosure tends to be more episodic or time-bound. Filing-level analyses further show that lifecycle outcomes vary with topic composition: greater concentrations of Labour Relations and Workplace Health and Safety are associated with higher rates of both introduction and abandonment, while Indigenous Community Engagement is associated with higher introduction but not abandonment. DEI Initiatives show no systematic association with either lifecycle outcome. Finally, institutional alignment is associated with greater disclosure continuity overall, but this relationship varies with topic composition. Greater Workplace Health and Safety concentration weakens the negative association between institutional alignment and lifecycle change, whereas greater Indigenous Community Engagement concentration strengthens it.

The study makes several contributions to the accounting literature. First, it advances research on social disclosure by moving beyond the analysis of isolated disclosure themes. Prior

studies have provided important insights into specific areas of social reporting, including diversity and inclusion, human capital, and community relations (Baker et al., 2024; Demers et al., 2024; Kent and Zunker, 2013; Liang et al., 2021; Soobaroyen and Mahadeo, 2016). This study extends this literature by examining how different social issues are organized within a broader voluntary disclosure portfolio. In doing so, it shifts attention from individual social topics to the structure of social visibility within corporate reporting.

Second, it contributes to research on institutional influences in voluntary disclosure. Prior accounting research suggests that regulatory pressures and institutional expectations can lead to more consistent and transparent corporate disclosures by offering firms "a presentation of their diverse audiences' perceptions" (Luft Mobus, 2005; Chelli et al., 2014, p. 307; Cho and Patten, 2007; Li and Yang, 2016). However, this perspective often implies a relatively uniform stabilizing effect, in which firms respond to institutional expectations by aligning their disclosures with prevailing norms. This study refines that view by showing that institutional alignment is associated with disclosure continuity and that this relationship varies with the composition of social disclosure.

Third, it contributes methodologically to research on disclosure dynamics by shifting the unit of temporal comparison from documents to sentences. Prior accounting research shows that year-over-year disclosure change matters, but it has generally examined change by comparing entire documents or disclosure sections across periods (Brown and Tucker, 2011; Peterson et al., 2015). Such approaches are useful for measuring overall modification or consistency, but they provide limited insight into which specific pieces of disclosure content are newly introduced, carried forward, or abandoned. This study addresses this limitation by classifying sentence-level disclosure content across adjacent reporting years and then aggregating those classifications to

examine topic- and filing-level lifecycle patterns. This design makes it possible to observe the substantive composition of disclosure change.

The rest of the paper is structured as follows. Section 2 develops the theoretical framing and research questions. Section 3 describes the data and methods. Section 4 presents the findings. Section 5 discusses the implications, contributions, limitations, and directions for future research.

2. Theoretical Framework

2.1 Visibility

Visibility provides the first theoretical anchor for understanding voluntary social disclosure. Accounting scholars have long argued that disclosure is not simply a neutral mechanism for transmitting information. Rather, disclosure participates in the construction of organizational reality by making some issues visible while leaving others less visible or entirely absent (Burchell et al., 1980; Hopwood, 1987; Preston et al., 1996; Neu et al., 1998; Carmona et al., 2002). Visibility is therefore not an inherent property of organizational issues but an outcome of accounting practices that determine what is presented, emphasized, and communicated to external audiences.

This perspective is particularly relevant in the context of voluntary social disclosure. Because much social disclosure remains voluntary, organizations retain considerable discretion regarding what information is communicated and how it is presented. Visibility within voluntary disclosure is therefore inherently selective. Firms may choose to emphasize particular social issues, stakeholder relationships, and organizational initiatives while devoting comparatively less attention to others. Prior social and environmental disclosure research consistently demonstrates such selectivity, showing that firms do not disclose all social issues equally (Cowen et al., 1987; Deegan and Rankin, 1996; Cho et al., 2010, 2012; Michelon et al., 2015). Voluntary disclosure

therefore reflects not a comprehensive account of organizational social performance but a particular configuration of issues that firms choose to make visible.

From this perspective, the key analytical question is not simply whether firms disclose social information, but which social issues achieve visibility within disclosure. Topic structure provides a useful lens for examining this distribution of visibility: dominant disclosure topics reveal the issues that firms repeatedly foreground in their communications and the forms of social activity that occupy the most prominent positions within voluntary disclosure.

2.2 Temporality

While visibility helps identify which social issues become prominent within voluntary disclosure, it provides only a snapshot of disclosure at a particular reporting moment. Understanding disclosure therefore requires attention not only to what is visible, but also to how visibility unfolds through time. My second theoretical anchor draws on the concept of temporality.

The importance of temporality has been increasingly recognized within accounting research. Prior studies argue that accounting practices do not simply record organizational activities occurring through time; they also organize those activities through temporal structures such as reporting periods, budgets, forecasts, and performance cycles (Klein Jr and Reilley, 2024). As Ezzamel and Robson (1995) argue, accounting therefore does more than document organizational activity. It creates the temporal framework through which activities can be interpreted over time. Consequently, accounting makes it possible to examine not only organizational change but also the continuity of organizational representations across successive reporting periods.

A central implication of this literature is that accounting draws attention to continuity as well as change. Because accounting operates through recurring reporting cycles, organizational

representations are repeatedly revisited and reassessed over time. Parker (2004) argues that temporal analysis requires attention not only to discontinuities and moments of change, but also to the forms of continuity that persist across successive periods. This insight is particularly important because continuity is often less visible than change. Organizational practices that are repeatedly reproduced across reporting cycles can become taken for granted, even though they play an important role in shaping how organizations are understood and represented over time. Examining temporality therefore involves understanding not only how organizational representations change, but also how they persist.

From this perspective, visible social issues have a temporal status. Some become durable elements of reporting, reappearing across successive filings and thereby acquiring continuity within reporting practice. Others remain episodic, appearing in response to particular circumstances but not becoming part of recurring disclosure practice. These differences matter because durability and transience imply different positions within the disclosure portfolio. Social issues that repeatedly reappear across reporting periods become recurring elements of the reporting portfolio and contribute to stable representations of organizational priorities. In contrast, issues that appear only intermittently may reflect shifting reporting attention. Temporality therefore extends the visibility perspective by directing attention not only to which social issues become visible, but also to the extent to which that visibility endures through time.

2.3 Institutional Expectations

If visible social issues differ in their temporal trajectories, the next question concerns the factors associated with these differences. Prior research has suggested that voluntary disclosure is produced within a broader institutional environment that helps define which issues are considered relevant, appropriate, and worthy of communication (Chelli et al., 2014; Cho and Patten, 2007;

Tilling and Tilt, 2010; Nazari et al., 2015; Li and Yang, 2016). My third theoretical anchor therefore draws on the concept of institutional expectations.

Early institutional theorists argued that organizations adopt structures and practices not solely because of their technical efficiency but because conformity with prevailing institutional norms confers social acceptance within their organizational field (Meyer and Rowan, 1977). Building on this foundation, DiMaggio and Powell (1983) identified three isomorphic mechanisms through which organizations respond to their institutional environment, including coercive pressures arising from formal rules and regulations, mimetic pressures arising from imitation under conditions of uncertainty, and normative pressures arising from professional standards and socially accepted practices.

The present study focuses primarily on normative institutional expectations. Unlike mandatory disclosure requirements, the voluntary social disclosures examined here are not subject to direct regulatory compulsion. Firms therefore retain substantial discretion regarding what information is reported and how it is presented. However, this discretion does not mean that disclosure choices are made in an institutional vacuum. Firms operate within an environment shaped by non-binding reporting frameworks, stock exchange guidance, professional norms, and widely recognized social responsibility principles. These arrangements provide reference points regarding the issues that are considered important, legitimate, and worthy of organizational attention (Chen and Roberts, 2010; Chelli et al., 2014). In situations characterized by discretion and ambiguity, such institutional expectations may be associated with how firms prioritize, communicate, and revisit particular social issues within voluntary disclosure.

From this perspective, institutional expectations provide a useful lens for examining variation in voluntary social disclosure. Different social issues may occupy different positions

within prevailing reporting frameworks and institutional guidance. This suggests that the broader normative environment in which firms operate may be related to how social narratives evolve over time.

Taken together, the concepts of visibility, temporality, and institutional expectations provide a framework for examining not only which social issues become prominent within voluntary disclosure, but also how those issues evolve through time and how they relate to broader institutional expectations. Guided by this framework, the study addresses the following research questions:

RQ1: What are the dominant voluntary social topics in Canadian MICs?

RQ2: How do disclosure lifecycle patterns vary across dominant voluntary social topics?

RQ3: What is the relationship between institutional alignment and disclosure lifecycle outcomes?

3. Data and Method

3.1 Sample Selection

I obtain MICs from Avantis AI, which provides Canadian disclosure documents in machine-readable text format. The initial corpus contains 48,399 MIC filings from 1997 to 2023. As shown in Table 3.1, I parse the filings into sentences using SpaCy and apply standard preprocessing procedures to remove non-English text, page numbers, headers, footers, very short sentences (fewer than 30 characters), digits, and non-alphanumeric characters. This procedure yields 1,450,748 parsed sentences from 47,219 filings.

Table 3.1 Sample Selection

	Filings	Sentences
Initial MIC filings (1997–2023)	48,399	
Less: excluded during text preprocessing	(1,180)	
Parsed sentences after preprocessing		1,450,748
Less: sentences not matching social keyword dictionary		(1,299,761)
Less: sentences matching governance/boilerplate exclusion criteria		(32,777)
Less: duplicate sentences		(2,630)
Less: mandatory board gender diversity disclosures		(18,097)
Less: filings with no remaining social sentences	(29,104)	
Less: filings without adjacent-year data	(5,341)	(16,593)
Lifecycle analysis sample	12,774	80,890

Note: This table reports the sample selection process for voluntary social disclosure sentences in Canadian management information circulars (MICs) from 1997 to 2023.

To identify social sentences, I adopt a dictionary-based approach commonly used in corporate disclosure research (Loughran and McDonald, 2011; Matsumoto et al., 2011). I begin with social keyword dictionaries from Baier et al. (2020), Lin et al. (2024), and Loughran et al. (2023), and refine them by reviewing high-frequency terms in the corpus. The final dictionary captures five broad areas of voluntary social disclosure: diversity and inclusion, labour standards, employee wellbeing, community, and human rights and CSR (see Appendix B, Panel A). Applying this dictionary yields 150,987 candidate sentences.

Because MICs contain many governance, compensation, and legal sentences that can be falsely captured by social keywords, I apply a two-stage exclusion procedure. First, I remove sentences appearing in executive compensation descriptions, clinical and medical contexts, financial statement content, meeting-related COVID boilerplate, committee mandate descriptions,

and other recurring non-social contexts. This step removes 32,777 sentences, after which deduplication removes a further 2,630 sentences that appear verbatim across multiple firm-years.[15]

Second, to ensure the sample contains only voluntary social disclosures, I exclude sentences fulfilling the mandatory board gender diversity disclosure requirements under Items 11–15 of Form 58-101F1, effective December 31, 2014.[16] I use regular expression patterns to identify sentences containing standard compliance language associated with these items, including references to the proportion of women on the board, written diversity policies, designated group representation statistics, and explanations for the non-adoption of women's representation targets. This step removes 18,097 sentences and leaves 97,483 voluntary social sentences from 18,115 filings across 3,516 firms. The full exclusion list is provided in Appendix B, Panel B.

For the lifecycle analysis, I retain firm-years with adjacent-year observations so that each sentence can be compared with the firm's social disclosure content in the previous or subsequent year. This produces two distinct lifecycle samples. The year *t* sample contains 80,890 sentences from 12,774 firm-year pairs across 2,408 firms, and is used to classify current-year sentences as either INTRODUCED or CARRIED_FORWARD. The year *t−1* sample contains 70,483 sentences from the same firm-year pairs, and is used to classify prior-year sentences as ABANDONED or retained.[17] This separation ensures that introduction and continuation decisions are estimated on

[15] Exact duplicate sentences are removed to avoid mechanically inflating the prominence of recurring boilerplate language. This is particularly relevant for less dominant themes that could otherwise appear disproportionately important because a small number of standardized sentences is repeated across firm-years. Results are substantively similar when duplicate sentences are retained.

[16] Items 11–15 were added to Form 58-101F1 effective December 31, 2014, requiring non-venture issuers to disclose: their written policies for identifying and nominating women directors (Item 11); the extent to which women's representation is considered in director and executive appointment processes (Items 12–13); measurable targets for women's representation, if adopted (Item 14); and the actual number and proportion of women on the board and in executive officer positions (Item 15).

[17] The year t-1 sample contains fewer sentences than the year t sample (70,483 vs. 80,890) because the lifecycle classification requires each firm-year to have an observable subsequent year. Firm-years at the end of the sample period (i.e., the most recent filing year for each firm) are included in the year t sample but have no subsequent year against which to assess abandonment, and are therefore excluded from the year t−1 sample. The asymmetry reflects

current-year content, while abandonment is estimated on prior-year content. All descriptive and regression analyses are conducted on the lifecycle sample. The regression sample is further restricted to firm-years with non-missing values on all financial control variables, yielding 8,591 firm-year observations from 1,377 firms for the PCT_INTRODUCED specification and 7,565 firm-year observations from 1,254 firms for the PCT_ABANDONED specification.

3.2 Topic Modeling

To identify the main voluntary social topics in MIC disclosures, I apply BERTopic (Grootendorst, 2022) to the year-*t* lifecycle sample of 80,890 voluntary social sentences. BERTopic is appropriate for this study because it combines transformer-based sentence embeddings with dimensionality reduction and density-based clustering, allowing topics to emerge from sentence-level semantic similarity rather than from a pre-specified topic structure.

I first encode all sentences using FinBERT-ESG (Huang et al., 2023), a financial-domain language model trained on ESG-related corporate text.[18] The model converts each sentence into a high-dimensional embedding that captures its semantic content. I then use UMAP (McInnes *et al.*, 2018) to reduce the embedding dimensionality to five components (n_neighbors = 15, min_dist = 0.0, cosine metric), followed by HDBSCAN clustering (Campello *et al.*, 2013) with a minimum cluster size of 150 and cluster selection method "eom." Sentences that do not belong to any coherent cluster are assigned to a noise cluster (topic -1).[19]

---

the forward-looking nature of the abandonment classification relative to the backward-looking nature of the introduction classification.

[18] Due to this capability, FinBERT has been increasingly applied in corporate disclosure research (Alias *et al*., 2023; Kim *et al*., 2023).

[19] One strength of HDBSCAN is that it can effectively detect outliers (i.e., sentence embedding with insufficient proximity to other points) and assign them to a special cluster labeled as -1. This enhances clustering robustness by filtering out irrelevant or noisy sentences, allowing for more focused thematic identification. Studies have demonstrated the efficacy of HDBSCAN for tasks like fraud detection based on behavioral patterns (Min *et al*., 2021) and theme identification in personal finance inquiries (Qiu *et al*., 2021).

The model identifies 71 topic clusters and a noise cluster. I focus the main analysis on the five largest interpretable social topics: Community and Philanthropy, Workplace Health and Safety, DEI Initiatives, Indigenous Community Engagement, and Labour Relations. Together, these five topics account for 24,058 sentences, representing approximately 30% of the voluntary social sentence sample. The remaining sentences are either assigned to smaller topics or to the HDBSCAN noise cluster. I label each topic through close reading of representative sentences located near the topic centroid.

3.3 Lifecycle Classification

To examine how voluntary social disclosures change over time, I classify sentences into lifecycle stages using year-over-year semantic similarity within the same firm. For each firm-year pair, I compute cosine similarity between FinBERT-ESG sentence embeddings in year t and year $t-1$.

For each sentence disclosed in year t, I identify its maximum similarity to any voluntary social sentence disclosed by the same firm in year $t-1$. If the maximum similarity is at or above 0.80, I classify the sentence as CARRIED_FORWARD, meaning that a semantically similar sentence appeared in the prior year. If the maximum similarity is below 0.80, I classify the sentence as INTRODUCED, meaning that the current-year sentence does not have a sufficiently similar prior-year counterpart.[20]

I apply the same logic in reverse to identify abandoned content. For each sentence disclosed in year $t-1$, I identify its maximum similarity to any voluntary social sentence disclosed by the same firm in year t. If the maximum similarity is below 0.80, I classify the sentence as

[20] I report robustness checks using alternative cosine thresholds of 0.75 and 0.85 in Appendix C. The classification patterns in the results are consistent across thresholds.

ABANDONED, meaning that the prior-year sentence no longer has a sufficiently similar current-year counterpart.

This design treats introduction/continuation and abandonment as related but distinct disclosure decisions. INTRODUCED and CARRIED_FORWARD partition the year *t* sample and therefore sum to 100% within that sample. ABANDONED is estimated separately on the year *t−1* sample because abandonment concerns prior-year content that does not recur in the current year. This partitioning addresses the concern that a single estimation sample would conflate decisions made at different points in the disclosure cycle. The year *t* lifecycle sample comprises 80,890 sentences from 12,774 adjacent firm-year pairs across 2,408 firms and is used to estimate introduction and continuation. The year *t−1* sample comprises 70,483 sentences from the same adjacent firm-year pairs and is used to estimate abandonment.

3.4 Analysis of Lifecycle Patterns

To address RQ2, I use two complementary approaches. First, I examine the substantive content associated with different lifecycle stages within each topic. Second, I conduct filing-level regression analyses to assess whether lifecycle outcomes vary systematically with topic composition. Together, these analyses provide both a qualitative and quantitative assessment of disclosure lifecycle patterns.

Topic-Stage Content Analysis

To examine the substantive content associated with disclosure lifecycle stages, I identify representative sentences for each topic-stage combination. For each of the five dominant social topics and each lifecycle stage, I select the sentence whose FinBERT-ESG embedding is closest to the corresponding topic-stage centroid. These representative sentences are used to characterize the

dominant subtopics associated with introduced, carried-forward, and abandoned disclosure. Appendix F reports illustrative examples for each topic-stage combination.

Baseline Regression Model

To assess whether lifecycle outcomes vary systematically with topic composition, I estimate filing-level regression models using PCT_INTRODUCED and PCT_ABANDONED as dependent variables.

PCT_INTRODUCED is calculated as the number of introduced sentences divided by the total number of voluntary social sentences in the year *t* filing. PCT_ABANDONED is calculated as the number of abandoned sentences divided by the total number of voluntary social sentences in the year $t-1$ filing. The analysis focuses on PCT_INTRODUCED and PCT_ABANDONED because PCT_CARRIED_FORWARD is mechanically determined by PCT_INTRODUCED within the year *t* sample. As introduced and carried-forward content sum to 100 percent, estimating a separate carried-forward model would not provide additional information.

The key explanatory variables are topic-share measures corresponding to the five dominant social topics identified in the topic-modeling analysis. Each topic share is calculated as the proportion of voluntary social sentences assigned to a given topic within a filing. These variables capture the relative prominence of different social issues within a firm's disclosure portfolio.

The models additionally include controls for profitability (ROA), leverage (LEVERAGE), market-to-book ratio (MTB), property, plant and equipment intensity (PPE), firm size (LOG_ASSETS), asset turnover (TURNOVER), capital expenditure intensity (CAPEX), and Big Four auditor status (AUDITOR_BIG4). Variable definitions are reported in Appendix E.

All models are estimated using post-LASSO OLS with firm fixed effects and firm-clustered standard errors. Post-LASSO OLS follows a two-step procedure. LASSO is first used to

select explanatory variables, after which coefficients are re-estimated using ordinary least squares on the selected model. The approach reduces overfitting and multicollinearity in settings with multiple candidate predictors but does not address endogeneity or support causal inference (Tibshirani, 1996). Variables retained by the LASSO procedure are reported in the regression tables, while omitted variables are indicated by "—".

The regression samples are constructed by merging disclosure variables with financial data. After listwise deletion of firm-years with missing control values, the PCT_INTRODUCED sample contains 8,591 firm-year observations from 1,377 firms, and the PCT_ABANDONED sample contains 7,565 firm-year observations from 1,254 firms.[21] Descriptive statistics for the regression samples are discussed in section 4.

3.5 Institutional Expectations and Lifecycle Outcomes

To address RQ3, I extend the baseline model by incorporating a measure of institutional expectation alignment and its interactions with topic composition.

Institutional alignment is measured using a 43-sentence reference corpus derived from widely recognized social reporting frameworks relevant to Canadian listed firms, including the TMX/TSX ESG Disclosure Primer, the UN Global Compact Principles, the GRI 400 Social Standards, and the Truth and Reconciliation Commission Calls to Action. Appendix D reports the full corpus and source references.

For each voluntary social sentence, I calculate its maximum cosine similarity to the reference corpus. For example, a MIC sentence stating that the firm monitors risks associated with "paying a minimum living wage, restricting maximum working hours, and guarding against child

[21] As a robustness check, I re-estimate all models on the full sample of 12,774 firm-years using mean imputation for missing financial controls. The main results are qualitatively unchanged.

labor and forced labor" is compared against all 43 reference sentences. Its highest similarity is obtained with the UN Global Compact principle stating that businesses should "uphold the elimination of all forms of forced and compulsory labour" and "the effective abolition of child labour." Because these sentences express closely related labour-rights concerns despite using different wording, the MIC sentence receives a high institutional alignment score. I then aggregate sentence-level similarity scores to the firm-year level by taking the mean similarity across all voluntary social sentences in the filing. The resulting measure, INST_ALIGNMENT, captures the extent to which a firm's disclosure portfolio resembles established institutional expectations regarding social reporting.

The extended specification includes INST_ALIGNMENT and topic-share-by-INST_ALIGNMENT interaction terms. These interactions test whether the association between topic composition and lifecycle outcomes varies with the degree of institutional alignment. All other model specifications remain identical to the baseline model.

4. Findings

4.1 Voluntary Social Topics

To address RQ1, I first identify the dominant voluntary social disclosure topics and then examine the content associated with different lifecycle stages within those topics. Table 3.2 reports the five largest voluntary social disclosure topics identified by BERTopic. Community and Philanthropy and Workplace Health and Safety are the two largest topics, accounting for 9.0% and 8.9% of voluntary social sentences, respectively. DEI Initiatives accounts for a further 7.6%, while Indigenous Community Engagement and Labour Relations represent smaller but substantively distinct portions of the corpus, at 2.9% and 1.4%, respectively.

Table 3.2 Top 5 Social Topics

Topic	N	%	Representative Sentence
Community and Philanthropy	7246	9.0%	At AirBoss, we are passionate about giving back to the community and participating in philanthropic causes that reflect the company's values, as evidenced by the involvement of many employees and executives in supporting charities and local communities. Corporate community involvement, such as employee participation in community fundraising programs, raised millions of dollars for more than [X] charities and community groups across Canada and assists in attracting and retaining employees.
Workplace Health and Safety	7234	8.9%	The project had an excellent safety record. No labour stoppage impacted construction activities, as the corporation maintained positive relations with its contractors and their employees. All subsidiary companies set comprehensive health and safety goals that measure performance and identify opportunities for improvement.
DEI Initiatives	6116	7.6%	Our company is committed to equal pay for equal work, including racial and gender pay equity, by compensating all employees fairly and equitably. Initiatives include investing in organizations that focus on social and racial justice, as well as supplier diversity programs.
Indigenous Community Engagement	2317	2.9%	This new company will enable us to build upon our Aboriginal partnerships and relationships across Canada and the North, positioning the company not only as an expert in building in remote communities but also as one with a demonstrated understanding of the cultural values and requirements of Aboriginal communities. We have awarded millions of dollars' worth of contracts to Indigenous communities and contractors engaged on the project.
Labour Relations	1,145	1.4%	Further relations with employees may be affected by changes in the scheme of labour relations that may be introduced by the relevant South African governmental authorities. In addition, Aritzia further identifies risks and opportunities through information gathered from our annual Aritzia Asks employee survey, as well as Aritzia’s monitoring of supplier labour rights performance.
Undominant Topics	30,775	38%	—
Noise	26,057	32.2%	—

Note: This table reports the five largest voluntary social topics identified through BERTopic. Percentages are calculated relative to the year t lifecycle sample of 80,890 voluntary social sentences. Representative sentences are selected based on proximity to the topic centroid and are used to illustrate the substantive content of each topic. Numerical values and dates are replaced with [X] during preprocessing to reduce identifying information. Undominant topics refer to smaller interpretable clusters not included in the main analysis, while noise refers to sentences assigned to the HDBSCAN noise cluster.

The representative sentences in Table 3.2 indicate that these topics capture different domains of social accountability. Community and Philanthropy centers on corporate giving, employee participation in community fundraising, and support for local communities. Workplace

Health and Safety is more operational in orientation, emphasizing safety records, contractor and employee relations, and formal health and safety goals. DEI Initiatives focuses on equal pay, racial and gender pay equity, social and racial justice, and supplier diversity. Indigenous Community Engagement is organized around Aboriginal and Indigenous partnerships, relationships with remote communities, cultural understanding, and contracts awarded to Indigenous communities and contractors. Labour Relations captures a narrower set of employment-related disclosures, including labour-relations schemes, employee surveys, and supplier labour rights monitoring.

4.2 Topic Content and Disclosure Lifecycle Dynamics

Topic-Stage Content Analysis

To address RQ2, I first examine the content associated with introduced, carried-forward, and abandoned disclosure within each topic. Table 3.3 summarizes the dominant subtopics for each topic-stage combination, while Appendix F provides representative sentences that illustrate the underlying disclosure content.

Table 3.3 Lifecycle Profiles of Dominant Voluntary Social Topics

	Introduced Content		Carried Forward Content		Abandoned Content	
Topic	Percent	Dominant Subtopic	Percent	Dominant Subtopic	Percent	Dominant Subtopic
Community & Philanthropy	16.6%	Community Outreach Initiatives	83.4%	Corporate Giving Programs	13.2%	Project-Specific Community Activities
Workplace Health & Safety	23.3%	Safety Performance Updates	76.7%	Safety Management Systems	18.1%	Past Safety Achievements
DEI Initiatives	14.1%	Diversity Recruitment Efforts	85.9%	Diversity and Inclusion Commitments	7.3%	One-Time Diversity Assessment Initiatives
Indigenous Community Engagement	14.1%	Reconciliation and Consultation Activities	85.9%	Indigenous Partnership Arrangements	10.7%	Consent and Project-Access Disputes
Labour Relations	38.6%	Collective Bargaining and Labour Disputes	61.4%	Employment Standards and Protections	29.4%	Workforce Management Initiatives

Note: This table summarizes the dominant lifecycle profile of each of the five largest voluntary social topics. Introduced and carried-forward percentages are calculated within the year t sample and therefore sum to 100 percent for each topic. Abandoned

percentages are calculated separately using the year t−1 sample because abandonment refers to prior-year sentences that do not recur in the subsequent filing. Dominant subtopics are identified through close reading of representative sentences for each topic-stage combination.

Across the five dominant topics, newly introduced disclosure is typically associated with new initiatives, stakeholder engagement activities, and emerging organizational priorities. For example, newly introduced community disclosures often focus on outreach programs and youth employment initiatives, while introduced Indigenous-related disclosures emphasize reconciliation objectives and consultation activities. Similarly, introduced DEI disclosure frequently centers on recruiting diverse candidates and expanding opportunities for underrepresented groups.

In contrast, carried-forward disclosure is more closely associated with recurring organizational programs, policies, and stakeholder relationships. Within Community and Philanthropy, carried-forward content is dominated by corporate giving programs and established community engagement frameworks. Health and safety disclosures are similarly characterized by references to formal safety programs and management systems. Indigenous-related carried-forward disclosure emphasizes long-term partnership arrangements and contracting relationships, while DEI disclosure focuses on ongoing organizational commitments to diversity and inclusion.

Abandoned disclosure exhibits a different profile. Rather than describing continuing organizational practices, abandoned content is more frequently tied to specific projects, events, assessments, or disputes. Examples include community development activities associated with particular projects, historical safety achievements, one-time diversity surveys, and Indigenous consent or project-access disputes. These disclosures appear more contingent on specific circumstances and therefore less likely to persist from one reporting period to the next.

Filing-Level Lifecycle Associations

Before turning to the regression analyses, Table 3.4 reports descriptive statistics for the regression variables. The mean value of PCT_INTRODUCED is 0.142, whereas the mean value of PCT_ABANDONED is 0.089. This difference indicates that firms introduce new voluntary social disclosure content more frequently than they abandon existing content. Alignment with institutional expectations exhibits a relatively high mean value of 0.780, suggesting that voluntary social disclosures are, on average, closely aligned with the institutional reference corpus.

Table 3.4 Descriptive Statistics

Variable	N	Mean	SD	P25	Median	P75	Min	Max
Panel A: PCT_INTRODUCED sample (N = 8,591 firm-years; 1,377 firms)								
PCT_INTRODUCED	8,591	0.142	0.240	0.000	0.000	0.200	0.000	1.000
INST_ALIGNMENT	8,591	0.780	0.057	0.750	0.788	0.818	0.346	0.930
$T_{Community_Philanthropy}$	8,591	0.063	0.155	0.000	0.000	0.033	0.000	1.000
T_{Health_Safety}	8,591	0.078	0.179	0.000	0.000	0.069	0.000	1.000
$T_{DEI_Initiatives}$	8,591	0.116	0.248	0.000	0.000	0.100	0.000	1.000
$T_{Indigenous_Engagement}$	8,591	0.014	0.064	0.000	0.000	0.000	0.000	1.000
$T_{Labour_Relations}$	8,591	0.018	0.101	0.000	0.000	0.000	0.000	1.000
ROA	8,591	-0.184	0.814	-0.128	-0.002	0.053	-8.154	0.606
LEVERAGE	8,591	0.223	0.301	0.000	0.138	0.333	0.000	2.051
MTB	8,591	3.292	9.086	0.911	1.356	2.359	0.157	99.295
PPE	8,591	0.529	0.461	0.111	0.503	0.827	0.000	2.997
LOG_ASSETS	8,591	5.779	2.783	3.900	5.777	7.648	-6.215	14.645
TURNOVER	8,591	0.522	0.641	0.020	0.294	0.764	0.000	3.172
CAPEX	8,591	0.245	9.297	0.009	0.045	0.117	0.000	817.500
AUDITOR_BIG4	8,591	0.775	0.417	1.000	1.000	1.000	0.000	1.000
Panel B: PCT_ABANDONED sample (N = 7,565 firm-years; 1,254 firms)								
PCT_ABANDONED	7,565	0.089	0.200	0.000	0.000	0.077	0.000	1.000
INST_ALIGNMENT	7,565	0.780	0.058	0.750	0.788	0.819	0.346	0.930
$T_{Community_Philanthropy}$	7,565	0.061	0.156	0.000	0.000	0.000	0.000	1.000
T_{Health_Safety}	7,565	0.080	0.186	0.000	0.000	0.062	0.000	1.000
$T_{DEI_Initiatives}$	7,565	0.041	0.119	0.000	0.000	0.000	0.000	1.000
$T_{Indigenous_Engagement}$	7,565	0.012	0.062	0.000	0.000	0.000	0.000	1.000
$T_{Labour_Relations}$	7,565	0.017	0.100	0.000	0.000	0.100	0.000	1.000
ROA	7,565	-0.180	0.812	-0.120	0.001	0.054	-8.154	0.606
LEVERAGE	7,565	0.218	0.293	0.000	0.134	0.327	0.000	2.051
MTB	7,565	3.342	9.015	0.923	1.387	2.400	0.157	99.295

PPE	7,565	0.537	0.474	0.112	0.506	0.832	0.000	2.997
LOG_ASSETS	7,565	5.788	2.779	3.910	5.782	7.655	-6.215	14.172
TURNOVER	7,565	0.535	0.653	0.025	0.298	0.794	0.000	3.172
CAPEX	7,565	0.245	9.472	0.010	0.046	0.118	-0.035	817.500
AUDITOR_BIG4	7,565	0.785	0.411	1.000	1.000	1.000	0.000	1.000

Note: This table reports descriptive statistics for the regression samples. Panel A reports the sample used for the PCT_INTRODUCED specification, and Panel B reports the sample used for the PCT_ABANDONED specification. PCT_INTRODUCED is the proportion of voluntary social sentences in a filing classified as newly introduced. PCT_ABANDONED is the proportion of prior-year voluntary social sentences that do not recur in the subsequent filing. INST_ALIGNMENT is the mean sentence-level cosine similarity between a firm-year's voluntary social disclosure and the institutional reference corpus. Topic variables measure the proportion of voluntary social sentences in a filing assigned to each topic. Variable definitions are provided in Appendix E.

The topic-share variables display substantial variation across firm-years, indicating meaningful differences in the composition of firms' voluntary social disclosure portfolios.[22] While some firm-years contain little or no disclosure associated with a given topic, others devote a substantial proportion of their voluntary social disclosure to particular topics. Similar variation is observed across the control variables, reflecting differences in firm size, profitability, leverage, asset structure, and reporting environments. Overall, the descriptive statistics indicate sufficient variation in both lifecycle outcomes and disclosure composition to support the regression analyses reported in the following sections.

Table 3.5 reports the filing-level post-LASSO OLS results relating topic shares to the rates of disclosure introduction and abandonment.

[22] Topic shares reflect within-firm-year concentration and may rank differently from the aggregate sentence counts in Table 3.2. A topic with fewer aggregate sentences may have a higher mean share if its content is concentrated in a smaller number of firm-years where it constitutes a large fraction of the social disclosure, whereas a topic with more aggregate sentences may have a lower mean share if its content is distributed broadly across many firm-years at low within-filing shares.

Table 3.5 Post-LASSO OLS Regression Results: Topic Lifecycle Profiles

	(1) PCT_INTRODUCED	(2) PCT_ABANDONED
$T_{Community_Philanthropy}$	0.0793**	0.0864***
	(0.0366)	(0.0333)
T_{Health_Safety}	0.1220***	0.0845***
	(0.0322)	(0.0263)
$T_{DEI_Initiatives}$	0.0452	-0.0578
	(0.0423)	(0.0353)
$T_{Indigenous_Engagement}$	0.1726**	-0.0962
	(0.0732)	(0.0885)
$T_{Labour_Relations}$	0.2606***	0.1725**
	(0.0777)	(0.0813)
ROA	—	-0.0024
		(0.0067)
LEVERAGE	—	—
MTB	0.0003	0.0004
	(0.0006)	(0.0006)
PPE	0.0345***	—
	(0.0110)	
LOG_ASSETS	-0.0034	-0.0018
	(0.0058)	(0.0047)
TURNOVER	0.0172	—
	(0.0112)	
CAPEX	—	—
AUDITOR_BIG4	0.2874***	—
	(0.0050)	
N (firm-years)	8,591	7,565

*Note: This table reports post-LASSO OLS regressions examining the association between topic composition and disclosure lifecycle outcomes. The dependent variables are PCT_INTRODUCED in Column (1) and PCT_ABANDONED in Column (2). Topic variables measure the proportion of voluntary social sentences in a filing assigned to each dominant topic. All models include firm fixed effects and firm-clustered standard errors. Standard errors are reported in parentheses. "—" indicates variables not selected by the LASSO procedure. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.*

The results indicate that filing-level disclosure dynamics vary systematically with topic composition. Greater concentration in Labour Relations is associated with significantly higher rates of both disclosure introduction ($\beta = 0.261$, $p < 0.01$) and disclosure abandonment ($\beta = 0.173$, $p < 0.05$). Workplace Health and Safety displays a similar pattern, with greater topic concentration associated with higher introduction ($\beta = 0.122$, $p < 0.01$) and abandonment ($\beta = 0.085$, $p < 0.01$).

Thus, firm-years in which Labour Relations or Workplace Health and Safety occupies a larger share of voluntary social disclosure tend to exhibit greater overall disclosure turnover.

Community and Philanthropy exhibits smaller positive associations with both lifecycle outcomes. Greater community-related topic share is associated with higher disclosure introduction ($\beta = 0.079$, $p < 0.05$) and abandonment ($\beta = 0.086$, $p < 0.01$). Firm-years placing greater relative emphasis on Community and Philanthropy therefore also tend to exhibit greater overall disclosure change, although the estimated associations are smaller than those observed for Labour Relations and Workplace Health and Safety.

The results for DEI Initiatives and Indigenous Community Engagement differ from the preceding topics. DEI topic share is not significantly associated with either introduction or abandonment, indicating that greater relative emphasis on DEI does not correspond to systematic differences in filing-level lifecycle outcomes. Indigenous Community Engagement, by contrast, is positively associated with disclosure introduction ($\beta = 0.173$, $p < 0.05$) but not with disclosure abandonment. Thus, firm-years with greater relative emphasis on Indigenous Community Engagement tend to exhibit a higher overall rate of newly introduced voluntary social disclosure, without a corresponding systematic difference in the overall abandonment rate.

Among the control variables, PPE is positively associated with disclosure introduction ($\beta = 0.035$, $p < 0.01$) and Big Four auditor status is also positively associated with disclosure introduction ($\beta = 0.287$, $p < 0.01$). The remaining control variables do not exhibit consistent associations across the specifications.

Taken together, the RQ2 results reveal variation in disclosure lifecycle patterns at two complementary levels. The topic-stage analysis shows that the substantive content associated with introduction, continuation, and abandonment differs systematically: newly introduced disclosure

tends to concern new developments and organizational responses, carried-forward disclosure tends to center on established systems and commitments, and abandoned disclosure tends to be more episodic or time-bound. The filing-level regressions further show that overall lifecycle dynamics vary with topic composition. Greater concentrations of Labour Relations and Workplace Health and Safety are associated with higher rates of both introduction and abandonment, Community and Philanthropy exhibits smaller positive associations with both outcomes, and Indigenous Community Engagement is associated with higher introduction but not abandonment. DEI Initiatives show no systematic association with either filing-level lifecycle outcome. These findings suggest that variation in disclosure lifecycle patterns is reflected both in the substantive content occupying different lifecycle stages and in the relationship between topic composition and overall disclosure dynamics.

4.3 Institutional Expectations and Lifecycle Outcomes

RQ3 examines the relationship between institutional alignment and disclosure lifecycle outcomes and whether this relationship varies with topic composition. Table 3.6 reports the extended models that add INST_ALIGNMENT and topic-by-INST_ALIGNMENT interaction terms to the baseline specifications.

Table 3.6 Post-LASSO OLS Regression Results: Alignment with Institutional Expectations as Moderator of Topic Lifecycle

	(1) PCT_INTRODUCED	(2) PCT_ABANDONED
Main effects		
INST_ALIGNMENT	-0.8992***	-0.9734***
	(0.1422)	(0.1317)
$T_{Community_Philanthropy}$	0.1333	0.2095
	(0.4438)	(0.4280)
T_{Health_Safety}	-0.4562	-0.5921**
	(0.3139)	(0.2657)
$T_{DEI_Initiatives}$	-0.2097	-0.3325
	(0.8160)	(0.6274)
$T_{Indigenous_Engagement}$	2.0159*	3.1201***

	(1.0449)	(1.0702)
$T_{Labour_Relations}$	-0.1281 (0.7025)	0.6045 (0.6872)
Moderating effects (Topic × INST_ALIGNMENT)		
$T_{Community_Philanthropy}$ × INST_ALIGNMENT	-0.0726 (0.5624)	-0.1632 (0.5438)
T_{Health_Safety} × INST_ALIGNMENT	0.7322* (0.4127)	0.8581** (0.3544)
$T_{DEI_Initiatives}$ × INST_ALIGNMENT	0.3576 (1.0186)	0.3830 (0.7801)
$T_{Indigenous_Engagement}$ × INST_ALIGNMENT	-2.1698* (1.3042)	-3.8484*** (1.3109)
$T_{Labour_Relations}$ × INST_ALIGNMENT	0.4600 (0.9349)	-0.6738 (0.9027)
Controls		
ROA	—	-0.0039 (0.0066)
LEVERAGE	—	-0.0069 (0.0151)
MTB	0.0004 (0.0006)	0.0006 (0.0006)
PPE	0.0333*** (0.0109)	—
LOG_ASSETS	-0.0001 (0.0057)	0.0019 (0.0046)
TURNOVER	0.0166 (0.0110)	—
CAPEX	—	—
AUDITOR_BIG4	0.2132*** (0.0169)	—
N (firm-years)	8,591	7,565

*Note: This table reports post-LASSO OLS regressions examining whether institutional alignment moderates the association between topic composition and disclosure lifecycle outcomes. The dependent variables are PCT_INTRODUCED in Column (1) and PCT_ABANDONED in Column (2). INST_ALIGNMENT measures the mean semantic similarity between a firm-year's voluntary social disclosure and the institutional reference corpus. Interaction terms capture whether the association between topic composition and lifecycle outcomes varies with institutional alignment. All models include firm fixed effects and firm-clustered standard errors. Standard errors are reported in parentheses. "—" indicates variables not selected by the LASSO procedure. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.*

The topic main effects show that several associations remain statistically significant in the extended models. Indigenous Community Engagement is positively associated with both

disclosure introduction ($\beta = 2.016$, $p < 0.10$) and disclosure abandonment ($\beta = 3.120$, $p < 0.01$), while Workplace Health and Safety is negatively associated with disclosure abandonment ($\beta = -0.592$, $p < 0.05$). The remaining topic main effects are not statistically significant. Because the models include interaction terms, these coefficients represent conditional associations when INST_ALIGNMENT equals zero and should not be interpreted as average topic effects.

INST_ALIGNMENT is negatively and significantly associated with both lifecycle outcomes. Higher level of institutional expectations alignment is associated with lower rates of disclosure introduction ($\beta = -0.899$, $p < 0.01$) and disclosure abandonment ($\beta = -0.973$, $p < 0.01$). This pattern indicates that firm-years whose voluntary social disclosure portfolios more closely resemble established institutional social-reporting expectations tend to exhibit greater disclosure continuity. As with the topic main effects, however, these coefficients are conditional in the presence of the interaction terms.

For Workplace Health and Safety, the interactions with INST_ALIGNMENT are positive for both introduction ($\beta = 0.732$, $p < 0.10$) and abandonment ($\beta = 0.858$, $p < 0.05$), indicating that greater topic concentration weakens the negative association between institutional alignment and both lifecycle outcomes. In contrast, the interactions for Indigenous Community Engagement are negative for both introduction ($\beta = -2.170$, $p < 0.10$) and abandonment ($\beta = -3.848$, $p < 0.01$), indicating that greater topic concentration strengthens this negative association. The interactions for the remaining topics are not statistically significant.

The control variables are broadly consistent with those reported in Table 3.5. PPE and Big Four auditor status remain positively associated with disclosure introduction, while the remaining controls do not exhibit consistent relationships across the specifications.

Taken together, the RQ3 results indicate that greater institutional alignment is associated with greater disclosure continuity, but this relationship is not uniform across topic composition. Greater emphasis on Workplace Health and Safety weakens the negative association between institutional alignment and both introduction and abandonment, whereas greater emphasis on Indigenous Community Engagement strengthens it. These results suggest that the relationship between institutional alignment and disclosure continuity varies with the substantive composition of the firm's voluntary social disclosure portfolio.

5. Discussion

This study examined voluntary social disclosure in MICs through the lenses of visibility, temporality, and institutional expectations. The findings show that disclosure lifecycle patterns vary both in the substantive content associated with different lifecycle stages and in the relationship between topic composition and filing-level disclosure dynamics. Institutional alignment is associated with greater disclosure continuity overall, although this association varies with the relative prominence of selected social topics. Taken together, these findings extend the analysis of voluntary disclosure from what firms make visible at a particular point in time to how that visibility is maintained and reconfigured across reporting periods.

A first implication concerns the expanding boundaries of corporate governance communication. MICs are formally designed to support shareholder voting and governance accountability, and prior research has consequently focused primarily on their communication of governance arrangements, executive compensation, and firm performance (Packard et al., 2023). Yet the dominant voluntary topics identified in this study concern employees, communities, Indigenous stakeholders, diversity, and workplace conditions. The presence of these topics suggests that governance documents serve a broader communicative role than their formal purpose

might imply. Stakeholder-oriented social narratives are not confined to dedicated sustainability reports but are incorporated into documents traditionally associated with shareholder accountability. This finding therefore points to an intersection between governance communication and broader social accountability.

A second implication concerns the relationship between visibility and durability. Disclosure research commonly examines visibility in terms of whether an issue appears in corporate reporting or how much attention it receives. The lifecycle perspective developed here adds a temporal dimension to this understanding. Information can become visible without necessarily remaining visible: some content enters the reporting portfolio only temporarily, while other content is repeatedly carried forward and becomes a more durable component of corporate reporting. Two social issues can therefore be similarly prominent in a particular filing while occupying very different temporal positions within the firm's disclosure practice.

This distinction matters because durability captures something that cross-sectional measures of disclosure presence or volume cannot observe. The topic-stage analysis shows that carried-forward disclosure tends to center on established systems, commitments, and stakeholder relationships, whereas introduced content more often concerns new developments and organizational responses and abandoned content tends to involve more episodic or time-bound matters. The lifecycle of disclosure therefore provides information about how social issues become incorporated into, maintained within, or removed from the reporting portfolio. Visibility is consequently not only a question of what firms choose to make visible, but also of what they continue to keep visible over time.

Institutional expectations add another layer to the temporal interpretation of disclosure. Greater institutional alignment is associated with greater disclosure continuity overall, but the

strength of this relationship varies with the composition of the disclosure portfolio. This pattern speaks to Oliver's (1991) argument that organizations are not simply passive recipients of institutional pressures but exercise agency in how they respond to them. Institutional alignment may therefore be associated with disclosure durability without corresponding to the same degree of continuity across different social reporting portfolios.

This study makes three contributions to the accounting literature. First, it contributes to research on disclosure dynamics by reconceptualizing how disclosure change is examined. Prior research on social disclosure dynamics has predominantly examined change through aggregate volume trends or event-driven comparisons — documenting, for instance, that disclosure expands following environmental incidents (Patten, 1992), contracts after reputational crises (Bellucci et al., 2021), or grows more formulaic over time (Lin et al., 2024). While informative, these approaches treat change as a unidimensional shift and offer limited insight into the micro-level processes through which social narratives are maintained and reconfigured across successive reporting cycles. This study advances a temporal perspective by introducing a lifecycle framework that disaggregates disclosure dynamics into three analytically distinct stages and captures whether visible content becomes durable within the reporting portfolio.

Second, the study contributes to institutional research on voluntary disclosure. Prior research has established that institutional pressures are associated with voluntary social disclosure, but has tended to treat these pressures as relatively uniform in their effects, with firms disclosing more, disclosing more consistently, or aligning their content with prevailing norms (Chelli et al., 2014; Cho and Patten, 2007; Tilling and Tilt, 2010). At the same time, research has shown that firms may nevertheless develop systematically different sustainability reporting characteristics through different stakeholder relationships and engagement strategies (Herremans et al., 2016).

The present study extends this line of research by showing that heterogeneity is reflected not only in reporting characteristics across firms but also in the temporal trajectories of specific disclosure topics within firms over time. This study refines this understanding by showing that the relationship between institutional expectations and disclosure lifecycle outcomes varies with the relative prominence of different topics in the disclosure portfolio.

Third, the study contributes to research on proxy statements and governance communication. Prior studies on proxy statement disclosures have largely focused on executive compensation (Balsam et al., 2016; Packard et al., 2023) as well as shareholder proposals on issues such as climate-related policies (Diaz-Rainey et al., 2024). However, voluntary CSR content in MICs remains largely unexplored. The study addresses this gap by shifting the focus to a set of social issues, including diversity, community engagement, and labor practices. In so doing, the study shows that MICs contain a broader stakeholder-facing layer of voluntary communication alongside their conventional governance and shareholder-voting functions.

While the study offers several insights, it is subject to certain limitations. First, although MICs provide a useful setting for examining shareholder-facing social disclosure, disclosure dynamics may differ in other reporting venues such as sustainability reports, annual reports, or integrated reports. Future research could examine whether similar patterns emerge across different disclosure contexts. Second, while the regulatory frameworks used in this study provide a theoretically grounded basis for measuring institutional alignment, alternative institutional reference points may produce different assessments of disclosure alignment and continuity. Future research could examine whether disclosure dynamics vary across different institutional frameworks, industries, and jurisdictions. Third, the lifecycle framework captures the persistence and turnover of disclosure content rather than the persistence and turnover of the underlying

organizational practices being described. Future research combining disclosure data with organizational outcomes could further clarify the relationship between disclosure continuity and organizational change.

Voluntary social disclosure unfolds across time in ways that aggregate metrics cannot capture. This study suggests that examining what disclosure content is introduced, sustained, and abandoned across successive reporting cycles provides insight into the temporal organization of corporate social narratives that cannot be obtained from any single filing. As demands for social accountability intensify, attending to patterns of retention and abandonment alongside the content visible at any given moment offers a more complete account of how social issues are represented within corporate reporting over time.

Chapter 4: Recurring and Outlet-Specific Social and Environmental Accounts in Corporate Reporting

Abstract

This study examines the recurrence and outlet-specificity of social and environmental (S&E) disclosure across connected corporate reporting outlets. I conceptualize S&E reporting as a connected documentary system in which S&E matters are rendered into accounts that vary in how they are positioned, reproduced, and made available for scrutiny. Drawing on 3,416 matched Annual Information Form (AIF) and Management Discussion and Analysis (MD&A) filing pairs from Canadian mining firms (1997–2024), I use FinBERT-ESG sentence embeddings and cosine similarity to classify S&E sentences into overlapping, AIF-specific, and MD&A-specific disclosures. Topic modeling and phrase co-occurrence network analysis examine how these positions differ in content and accountability logics, and a fractional logistic regression examines firm-level and temporal variation in cross-outlet recurrence. The findings show that disclosure distribution is highly selective. Overlapping disclosures concentrate in formalized accounts of environmental assessment, closure obligations, greenhouse gas regulation, and workplace safety, organized through a regulatory-compliance logic. AIF-specific disclosures engage more situated concerns, including project-level impact assessment, community relations, Indigenous land rights, and emissions targets, organized through a rights-and-responsibility logic. MD&A-specific disclosures focus on implementation progress, operational outcomes, and managerial estimates, and are organized through a performance-and-estimation logic. Cross-outlet recurrence has increased over time; it is positively associated with MD&A S&E disclosure volume and negatively associated with firm size, AIF filing length, and asset intensity. The study contributes to social and environmental accounting research by showing that disclosure selectivity operates not only through what firms disclose, but also through how particular accounts are positioned and reproduced across connected reporting documents.

Keywords

Social and environmental disclosure; disclosure distribution; documentary recurrence; accountability; mining industry

1. Introduction

Social and environmental (S&E) disclosure is often studied as a question of presence: whether firms disclose, how much they disclose, and what content they include or omit (Deegan et al., 2002; Patten, 2002; Michelon et al., 2020). This framing has generated important insights into the selectivity and partiality of corporate S&E reporting (Gray, 2007). Yet it leaves less visible another dimension of disclosure: its documentary location. In contemporary corporate reporting regimes, firms issue multiple connected filings, each with its own regulatory purpose, generic conventions, and intended audience. Together, these filings constitute a documentary architecture through which corporate accounts are organized and made available for scrutiny. Within this architecture, the same S&E matter does not necessarily appear in the same form across every outlet. Some accounts are rendered in forms that recur across multiple filings, while others remain confined to a particular document. The question of what firms disclose is therefore inseparable from a question of location: where particular S&E accounts are positioned, and what that positioning implies for how visible, comparable, and accountable those accounts become.

To explore this dimension of disclosure, I draw on three related bodies of work. The first conceptualizes corporate reporting as a calculative and representational practice through which organizational activities, relations, and consequences are rendered into particular forms of account (Burchell et al., 1980; Hopwood, 1987; Miller and Rose, 1990). From this perspective, the forms that reporting can accommodate are never neutral; making S&E matters reportable always involves translation, simplification, and classification that shape which aspects of a firm's social and environmental conduct can be expressed and where they can appear. The second draws on the concepts of genre and inscription to examine how such forms become located across documents. Regulatory filings are distinct reporting genres that carry different expectations about purpose,

audience, and presentation, shaping what can be said within each document and what counts as appropriate disclosure (Yates and Orlikowski, 1992). Whether a particular S&E account recurs across that generic boundary depends on whether it has been rendered in a form stable enough to be recognized and reproduced across documents, a process theorized through the notion of inscription and mobility (Latour, 1987; Robson, 1992). The third draws on accountability research to consider what these documentary positions make possible. If accounts provide reasons for conduct and make organizational actions available for scrutiny, then where an S&E account is located within the reporting system may reflect the kind of accountability it can support, including who can see it, compare it, and call it to question (Roberts and Scapens, 1985; Gray et al., 1997; Messner, 2009).

These theoretical perspectives inform three research questions. The first asks what S&E content characterizes overlapping and outlet-specific disclosures (RQ1). The second investigates what accountability logics underlie overlapping and outlet-specific S&E content (RQ2). The third examines what firm and temporal characteristics are associated with the extent of S&E disclosure recurrence across AIF and MD&A filings (RQ3).

The Canadian mining sector provides a useful setting for examining this issue. Mining firms face extensive environmental and social obligations relating to land reclamation, water stewardship, tailings management, greenhouse gas emissions, biodiversity impacts, Indigenous consultation and consent, and long-term ecological monitoring (Campbell, 2012; Lauwo et al., 2016). These obligations are embedded in project geographies, regulatory jurisdictions, and community relationships that resist easy standardization. At the same time, firms must give these matters a reportable form within a securities disclosure system oriented toward investor relevance, financial materiality, and decision-usefulness (Mutti et al., 2012; Jenkins and Yakovleva, 2006).

The tension between the situated complexity of mining's S&E consequences and the categories available within securities reporting makes this sector especially useful for examining how S&E accounts are distributed across corporate reporting outlets.

I examine this reporting architecture through two core securities filings: the AIF and MD&A. Both documents are filed within the same reporting cycle under Canada's continuous disclosure regime, but they organize corporate information in different ways (Ontario Securities Commission, 2015). The AIF provides a broader account of the firm's business, properties, risks, regulatory environment, and operating context, whereas the MD&A focuses on management's discussion of financial performance, liquidity, risks, and current operations. This makes the AIF and MD&A a useful pair for examining S&E disclosure distribution: they are sufficiently connected to contain related S&E accounts, but sufficiently distinct to show whether those accounts recur across filings or remain specific to a particular reporting outlet.

The empirical analysis is based on 3,416 matched AIF–MD&A firm-years filed by Canadian mining firms between 1997 and 2024, in which both filings contain at least one identifiable S&E sentence. I identify S&E sentences in both filings using a custom dictionary and use FinBERT-ESG sentence embeddings and cosine similarity to classify sentences into three categories: overlapping disclosures, which appear in substantively similar form across both filings; AIF-specific disclosures, which remain located within the AIF; and MD&A-specific disclosures, which appear only in the MD&A. Topic modelling, phrase co-occurrence network analysis, and regression analysis are then used to examine the content, accountability logics, and temporal and organizational variation associated with these documentary positions.

The findings show that S&E disclosures are unevenly distributed across the AIF–MD&A reporting system. Overlapping disclosures are concentrated in relatively formalized accounts of

environmental assessment, project location, closure requirements, greenhouse gas regulation, and workplace safety. AIF-specific disclosures are more heterogeneous and project-specific, emphasizing project-level impact assessment, community relations, water and waste management, emissions targets, and Indigenous land-rights contexts. MD&A-specific disclosures exhibit a third pattern, focusing on implementation progress, operational outcomes, managerial estimates, reclamation liabilities, employment outcomes, and water management measures. The phrase co-occurrence analysis further shows that these three documentary positions are associated with different account-giving logics. Overlapping disclosures are organized primarily through a regulatory-compliance logic, AIF-specific disclosures through a rights-and-responsibility logic, and MD&A-specific disclosures through a performance-and-estimation logic. Finally, cross-outlet recurrence became more prevalent over time but remained selective. Firms with more extensive MD&A S&E disclosure exhibit higher recurrence, whereas larger firms, longer AIF filings, and more asset-intensive firms exhibit lower recurrence, suggesting that greater disclosure breadth and operational complexity are associated with more outlet-specific disclosure rather than greater cross-filing repetition.

The study makes three contributions. First, the study contributes to accounting research by conceptualizing corporate reporting as a connected reporting system rather than a collection of discrete documents. Prior disclosure research has typically examined individual reports or compared disclosure attributes across firms and time periods (Li, 2008; Frankel et al., 2010; Brown and Tucker, 2011; Huang et al., 2014; Brown and Knechel, 2016; Dyer et al., 2017). While this work has generated important insights into disclosure practices, it provides a more limited understanding of how related disclosures are organized across multiple reporting outlets. This study addresses that limitation by examining how S&E accounts are distributed across connected

filings within the same reporting cycle. In doing so, it introduces documentary position and cross-outlet recurrence as concepts for examining how reporting systems organize different forms of disclosure. The contribution is therefore to shift attention from individual reports to the relationships among reports that collectively constitute corporate reporting.

Second, the study contributes to social and environmental accounting (SEA) research by providing a more differentiated understanding of disclosure selectivity. Prior SEA research has demonstrated that corporate disclosures are selective representations of organizational activities and impacts rather than comprehensive accounts of corporate conduct (Roberts, 1992; Neu et al., 1998; Cho and Patten, 2007; Gray, 2007; Cho et al., 2015). However, discussions of selectivity have often focused on the distinction between disclosed and undisclosed information. The findings of this study suggest that selectivity may also operate within the disclosed domain itself. Environmental and social matters do not appear to be rendered reportable in the same way. Some forms of S&E disclosure become more readily standardized, reproduced, and circulated within corporate reporting, whereas others remain more closely tied to particular projects, stakeholder relationships, and organizational contexts. Selectivity therefore concerns not only whether information is disclosed, but also the forms through which different S&E matters become stabilized and sustained within reporting practice.

Third, the study contributes by making disclosure distribution observable at the level of the individual sentence. Prior S&E disclosure research has compared disclosure attributes across firms or years (Brown and Tucker, 2011; Bochkay et al., 2023; De Franco et al., 2020). Where cross-document comparisons have been made, they have generally relied on aggregate document-level measures of similarity, which can establish that two documents differ but cannot reveal where within those documents the differences lie, or what kind of S&E content is implicated. This study

instead compares AIF and MD&A disclosures sentence by sentence, allowing each S&E sentence to be located within the reporting system as either cross-outlet or outlet-specific. This classification shifts the unit of analysis from the document to the individual sentence, making it possible to examine whether cross-outlet and outlet-specific sentences differ in their topic composition, phrase associations, and firm-year patterns.

The remainder of the paper is organized as follows. Section 2 develops the theoretical framing. Section 3 describes the data and method. Section 4 presents the findings and section 5 discusses the study's contributions, limitations, and avenues for future research.

2. Theoretical Framing

2.1 Corporate Reporting as a Calculative Practice

A foundational premise of this study is that corporate reporting does not simply transmit pre-existing organizational realities to external audiences. It is a calculative and representational practice through which organizational activities, relations, and effects are made visible, classified, and positioned in particular ways (Burchell et al., 1980; Hopwood, 1987; Miller and Rose, 1990; Miller, 1998).

Accounting classifications are therefore not neutral labels attached to an already-formed reality. They participate in forming the very objects that reporting presents (Carmona et al., 2002; Miller and O'Leary, 1994; Preston et al., 1997; Neu, 2000; Neu and Graham, 2004; Neu et al., 2026). Preston et al. (1997, p. 161), for example, show how the U.S. medicare payment system for elderly patients translated diverse medical needs into standardized diagnostic categories and fixed reimbursement rates, illustrating how accounting classifications can turn complex social needs into administrative and budgetary objects. Neu and Graham (2004, p. 579) show how accounting techniques in the Canadian Indian Department translated colonial policy into administrative

practice, rendering Indigenous peoples governable through financial controls and numerical representations. These studies show that accounting techniques can transform complex social and organizational matters into differentially reportable objects, making some administratively visible while rendering others marginal or difficult to express within existing reporting categories. This argument is directly relevant to S&E disclosure research because corporate disclosure is one of the key sites where environmental and social matters are translated into such reportable forms.

In SEA research, this insight resonates with studies showing that corporate reports do not simply mirror S&E impacts. Firms selectively organize S&E information by emphasizing some matters, marginalizing others, and framing organizational responsibility through particular textual and visual forms (Boiral, 2013; Cho et al., 2015; Michelon et al., 2015; Liu et al., 2025). This is especially important in the mining sector, where S&E matters are especially heterogeneous, spanning issues such as operational processes, regulatory obligations, environmental effects, community relations, remediation responsibilities, and long-term ecological consequences (Campbell, 2012, p. 139; Lauwo et al., 2016, p. 1042). Mining firms' reports cannot reproduce this complexity as an unstructured social or environmental reality. Disclosure necessarily involves selection, simplification, and translation into reportable form.

2.2 Inter-Documentary Ordering of S&E Disclosure

If corporate disclosure translates heterogeneous S&E matters into reportable forms, a further question is whether those forms remain tied to a particular report or recur across related reporting documents. This question is important because corporate reports are not interchangeable containers of information. They are genres: recurring communicative forms that carry expectations about purpose, audience, content, and use (Yates and Orlikowski, 1992, p. 304; 2002, p. 14). These expectations inform how information is selected, arranged, and made relevant within a particular

document. AIFs and MD&As both form part of securities reporting, but as distinct reporting genres, they may position S&E accounts differently. As a result, whether an S&E account appears across both filings or remains confined to a single reporting outlet is not merely a formatting choice. It reflects whether that account has been rendered in a form that can be accommodated within the conventions of multiple reporting genres.

The notion of inscription helps explain this process. Latour's notion of "immutable and combinable mobiles" highlights how inscriptions retain enough stability to be carried across contexts and made available for comparison and combination (Latour, 1987, p. 227). Robson (1992, p. 694) brings this insight into accounting, arguing that the accounting reports operate through inscription: "company accounts inscribe productive processes, labour, transactions, machines, buildings, cash flows, and loans, dispersed spatially and temporally but given a form by the text." Inscriptions thus include not only numbers, but also written texts, tables, charts, lists, and other material representations through which particular settings are made available for action (Qu and Cooper, 2011). Applied to S&E disclosure, this perspective suggests that narrative accounts can give S&E matters a textual form through which they can be recognized, reproduced, and compared across documents. Cross-document disclosure location is therefore theoretically significant because it makes visible which S&E accounts recur across connected filings and which remain tied to a particular documentary setting.

Because AIFs and MD&As are filed within the same reporting cycle, this study does not assume a direction of movement from one document to the other. I therefore use recurrence to refer to substantively similar S&E accounts appearing across both filings, rather than to imply that disclosure content originates in one filing and is transferred to the other. This framework suggests three theoretically distinct positions within the AIF–MD&A reporting system. Overlapping

disclosures appear in substantively similar form across both filings, indicating that certain S&E accounts have been stabilized in forms that recur across reporting genres. AIF-specific disclosures appear only in the AIF, suggesting that these accounts remain tied to the AIF and do not recur in the corresponding MD&A. MD&A-specific disclosures appear only in the MD&A, reflecting S&E accounts that are foregrounded within the management narrative but do not recur in the corresponding AIF. The latter two are referred to collectively as outlet-specific disclosures because they remain associated with a particular reporting outlet rather than recurring across the reporting system. These three positions are not assumed to be equivalent in their account-giving capacities, and their content, accountability logics, and temporal distribution are therefore treated as empirical questions.

2.3 Accountability Relations and Organizing Logics of S&E Disclosure

The preceding sub-sections establish that S&E disclosure is not simply a matter of reporting content, but a calculative practice through which heterogeneous S&E matters are translated into reportable forms and positioned across documentary locations. Building on this, this sub-section turns to the accountability relations through which disclosures become accounts.

Accounting does not organize information in the abstract. It operates within systems of accountability, in which accounts provide reasons for conduct and establish the basis on which actions are evaluated and responsibility is assigned (Roberts and Scapens, 1985, pp. 447-449; Roberts, 1991, p. 355). Accountability therefore concerns not only the provision of information, but also the conditions under which that information carries obligation (Gray et al., 1997, p. 334). S&E disclosures should thus be examined not merely as statements about social and environmental topics, but as accounts that organize responsibility, relevance, and answerability in particular ways.

I use the term organizing logics of disclosure to capture this form of account-giving. The term refers to the recurring ways in which disclosures connect matters, assign relevance, invoke responsibility, and establish the basis on which an S&E matter can be reported, evaluated, or questioned. It therefore directs attention beyond disclosure topics to the reasoning through which a matter is presented as, for example, a risk, a compliance obligation, a stakeholder relation, or an environmental consequence. These are not assumed empirical categories. Rather, they indicate the kinds of account-giving forms through which S&E matters may be made answerable within corporate reporting.

This account-giving perspective also requires caution. Prior SEA research has reminded me that corporate reports may expand the volume of S&E information while leaving important questions of stakeholder voice, answerability, and organizational impact unresolved (Cho and Patten, 2007, p. 646; Neu et al., 1998, p. 279; Deegan and Rankin, 1996, pp. 58-59). It would therefore be inappropriate to assume in advance that overlapping disclosures are better, more complete, or more accountable, or that outlet-specific disclosures are residual and less accountable. Instead, overlap and outlet-specificity are treated as empirical positions within the AIF–MD&A reporting system. The task is to examine what forms of account-giving these positions support.

Furthermore, this accountability perspective suggests that cross-outlet overlap should not be treated as a static textual feature. Accountability relations are historically and organizationally situated rather than fixed (Roberts, 1992, p. 609). If S&E matters are made reportable, positioned across documents, and organized through particular account-giving logics, then the extent of recurrence may vary as reporting expectations, disclosure routines, and firm-level reporting capacities develop. The analysis therefore needs to examine not only what recurs and how recurring content is organized, but also whether recurrence varies across time and firms.

Taken together, this theoretical framing shifts attention from S&E disclosure as information contained in individual reports to the ways in which S&E matters are translated, located, and organized across a connected reporting system. Accordingly, the study asks:

RQ1: What S&E content characterizes overlapping and outlet-specific disclosures?

RQ2: What accountability logics are associated with overlapping and outlet-specific S&E content?

RQ3: What firm and temporal characteristics are associated with the extent of S&E disclosure recurrence across AIF and MD&A filings?

3. Data and Method

3.1 Sample

The initial filing sample is obtained from Avantis.ai, a provider of Canadian corporate filings. The dataset contains 18,753 MD&A filings and 7,750 AIF filings issued by Canadian mining firms between 1997 and 2024. To facilitate direct comparison across reporting outlets, I identify firm-year observations for which both AIF and MD&A filings are available, yielding an initial matched sample of 3,584 AIF–MD&A pairs.

To identify S&E disclosures, I construct a custom S&E dictionary drawing on terminology from the Global Reporting Initiative standards for the mining sector, Sustainability Accounting Standards Board standards, the Task Force on Climate-related Financial Disclosures, and prior disclosure studies (Lin et al., 2024; Cannon et al., 2020). The dictionary captures six categories of S&E issues relevant to Canadian mining firms: climate change, ecosystem, pollution and waste, stakeholder and community relations, occupational health and safety, and mining-specific environmental obligations. To minimize over-capture of non-S&E content, dictionary entries are predominantly multi-word phrases (bigrams or trigrams), and terms associated with generic

governance, financial reporting, or industrial operations without a direct S&E referent are excluded. Single-word entries are retained only where the term is sufficiently specific to a mining S&E context to minimize false positives (e.g., biodiversity, silicosis). The final dictionary contains 359 phrases and is reported in Appendix G.[23]

Applying the dictionary to the matched filings yields 122,341 S&E sentences from AIF filings and 33,006 S&E sentences from MD&A filings. However, not all matched firm-years yield S&E sentences in both filings. Firm-years in which the AIF contains no identifiable S&E sentences or the MD&A contains no identifiable S&E sentences are excluded from the cross-outlet matching procedure described in Section 3.2, as no comparison is possible when one side of the pair is empty. This reduces the analytical sample to 3,416 firm-years in which both the AIF and MD&A contain at least one S&E sentence. All sentence counts and analyses reported below are based on this analytical sample.

3.2 Identifying Overlapping and Outlet-Specific Disclosures

Following the theoretical framework developed in Section 2, I classify S&E sentences into three categories: overlapping disclosures, which appear in substantively similar form across both the AIF and MD&A of the same firm-year; AIF-specific disclosures, which appear only in the AIF; and MD&A-specific disclosures, which appear only in the MD&A. AIF-specific and MD&A-specific disclosures are referred to collectively as outlet-specific.

All S&E sentences are first encoded using FinBERT-ESG, a transformer-based language model pre-trained on financial and ESG text (Huang et al., 2023). This domain-specific model is

[23] To assess the accuracy of the S&E sentence extraction procedure, I randomly sampled 150 sentences from the raw filing corpus—75 classified as S&E by the extraction rules and 75 classified as non-S&E—and independently coded each sentence as genuinely S&E-relevant or not. The extraction rules achieve a precision of 93.3% and a recall of 93.3%, with consistent performance across AIF and MD&A filings (AIF: 93.5%/93.5%; MD&A: 92.3%/92.3%), indicating that the procedure identifies S&E sentences with high accuracy and limited over- or under-capture.

preferred over general-purpose sentence transformers because S&E disclosure in mining filings contains specialized terminology whose semantic relationships are better captured by a model trained on financial and sustainability texts.

I then calculate cosine similarity between the FinBERT-ESG embeddings of AIF and MD&A sentences within each matched firm-year. To identify overlapping sentences, I apply a bipartite matching procedure. Within each firm-year, I first identify all AIF-MD&A sentence pairs with a cosine similarity exceeding 0.80. I then assign matches in descending order of similarity: the highest-similarity pair is matched first, and both sentences are removed from the pool of available candidates. This continues until no unmatched pairs above the threshold remain. AIF sentence that are not matched are classified as AIF-specific; MD&A sentences that are not matched are classified as MD&A-specific.[24]

This procedure yields 27,610 overlapping sentence pairs, 94,731 AIF-specific sentences, and 5,396 MD&A-specific sentences across 3,416 matched firm-years. By construction, the number of overlapping sentences is identical when counted from the AIF side and from the MD&A side, as each match involves exactly one sentence from each filing.

3.3 Topic Analysis

To characterize the S&E content of each disclosure category, I apply BERTopic (Grootendorst, 2022), a topic modelling approach that combines contextualized embeddings, UMAP dimensionality reduction, HDBSCAN density-based clustering, and class-based TF-IDF to generate interpretable semantic topics. Rather than using BERTopic's default embedding model,

[24] This threshold is intended to capture disclosures that convey substantively similar information despite variations in wording or sentence structure. Following prior accounting research (Demers et al., 2024), which treats a cosine similarity score of 0.80 as indicating high textual similarity, I classify sentence pairs with similarity scores of 0.80 or above as overlapping disclosures. As a robustness check, I repeat the matching procedure using cosine similarity thresholds of 0.75 and 0.85. The resulting recurrence patterns remain qualitatively similar to those obtained using the primary threshold of 0.80.

I supply the pre-computed FinBERT-ESG embeddings described in Section 3.2, ensuring methodological consistency between the matching and topic modelling procedures.

Topic modelling is conducted separately for each of the three disclosure categories to enable direct comparison of their thematic composition. For each group, I report the number of topics identified, the proportion of sentences assigned to the noise cluster (i.e., topic -1, representing sentences that do not cluster into any coherent topic), and the share of sentences covered by the five most prevalent topics. These statistics provide a basis for assessing the coherence and coverage of the topic structure.

For interpretive clarity, I focus on the five most prevalent non-noise topics within each group. For each topic, I review the representative sentences identified by the BERTopic model and report two representative sentences to illustrate the S&E content captured. I then assign an interpretive label summarizing the substantive content of the cluster.

To complement the separate topic models, I also run a pooled BERTopic model on all three disclosure groups combined and compare the distribution of overlapping, AIF-specific, and MD&A-specific sentences across shared topics. This pooled analysis, reported in Appendix H, provides an alternative comparison that does not depend on the alignment of separately estimated topic structures.

3.4 Phrase Co-occurrence Network Analysis

To examine the accountability logics associated with each disclosure category, I conduct a phrase co-occurrence network analysis. In this study, accountability logic is not treated as a directly observable attribute of a sentence. Rather, it is inferred from the recurring textual associations through which firms make S&E matters reportable, relevant, and open to evaluation. The analysis therefore asks whether overlapping, AIF-specific, and MD&A-specific disclosures

systematically connect S&E matters to different kinds of concepts, obligations, actors, consequences, and organizational arrangements. Phrase co-occurrence analysis is well suited to this purpose because it identifies which concepts are repeatedly brought together within the same disclosure sentence, allowing the analysis to move beyond topic labels and examine the forms of account-giving through which S&E matters are organized (Carley, 1993).[25]

I begin by extracting bigrams from the combined corpus of AIF and MD&A disclosures using a minimum document frequency threshold of 100 sentences and a maximum document frequency threshold of 5% of sentences. These thresholds ensure that retained phrases are sufficiently common to represent recurring disclosure patterns while excluding phrases that appear so frequently that they provide little discriminatory value.

Separate co-occurrence networks are then constructed for overlapping, AIF-specific, and MD&A-specific disclosures. Within each network, phrases are represented as nodes and two phrases are connected when they appear within the same sentence. Edge weights are based on the number of distinct firms in which a phrase pair co-occurs rather than the raw frequency of occurrence. This firm-level weighting reduces the influence of repeated boilerplate language within individual firms and emphasizes associations that recur across the reporting population.

To focus on recurring disclosure patterns rather than firm-specific language, I retain only phrase pairs that co-occur in at least 15 distinct firms for the overlapping and AIF-specific networks and at least 8 distinct firms for the MD&A-specific network. The higher threshold for the former groups reflects their substantially larger corpus size, while the lower threshold for the

[25] In accounting research, network analysis has been used to study both social and textual relations, such as auditor independence, CEO connections, investor relations networks, citation networks, and thematic associations in accounting texts (Wakefield, 2008; Neu et al., 2022; Bianchi et al., 2023).

MD&A-specific group accommodates its smaller number of sentences while still requiring that retained associations recur across multiple firms.

For each network, I report overall structural characteristics, including the number of nodes, number of edges, total edge weight, and network concentration. I then report the five strongest phrase-pair connections in each disclosure category, together with representative disclosure sentences. The strength of a phrase-pair connection is measured by its edge weight, defined as the number of distinct firms in which the two phrases co-occur within the same sentence. The analysis focuses on how S&E matters are repeatedly associated with particular concepts, obligations, actors, and organizational concerns. These recurring associations are interpreted as indicative of the accountability logics through which different forms of S&E disclosure are organized and made meaningful within the reporting system.

3.5 Cross-Outlet Recurrence Across Time and Firms

To examine how S&E disclosure recurrence varies across time and firms, I construct a filing-level recurrence ratio for each AIF filing, defined as the number of overlapping sentences divided by the total number of AIF S&E sentences in the filing. A higher ratio indicates that a greater proportion of the AIF's S&E content recurs in the corresponding MD&A.[26]

To examine temporal variation in recurrence, I aggregate the filing-level recurrence ratios by fiscal year and plot the annual mean alongside the number of firm-year observations per year. Reporting the annual observation count allows assessment of whether year-to-year fluctuations

[26] The recurrence ratio is constructed from the AIF side because the AIF is the larger and more comprehensive S&E filing in the AIF–MD&A pair, containing substantially more S&E sentences than the corresponding MD&A on average. A ratio constructed from the MD&A side would be mechanically higher for most firm-years, because the smaller denominator means that even a modest number of matched sentences produces a high ratio. This ceiling effect would compress variation across firms and reduce the sensitivity of the measure to differences in disclosure strategy.

reflect changes in disclosure behavior or shifts in sample composition, particularly in the early years of the sample period, where coverage is sparse.

To examine cross-firm variation, I use the filing-level recurrence ratio (RECURRENCE_RATIO) as the dependent variable in a regression analysis. Because the recurrence ratio is bounded between zero and one, the primary specification uses fractional logistic regression (Papke and Wooldridge, 1996), which is appropriate for proportional outcomes and avoids the boundary violations that can arise with ordinary least squares.

The independent variables capture several dimensions of firm characteristics that may be associated with variation in cross-outlet disclosure recurrence. LOG_ASSETS measures firm size, ROA captures profitability, and LEVERAGE reflects financing structure. TURNOVER, PPE, and CAPEX capture different aspects of firms' operating and investment characteristics. LOG_PAGES measures AIF filing length and serves as a proxy for disclosure breadth. I also include the logged number of MD&A S&E sentences (LOG_MDA_SE_COUNT) to control for the size of the MD&A S&E corpus: firms with more MD&A S&E sentences mechanically provide more potential matching candidates, which may inflate the observed recurrence ratio independently of systematic disclosure practices. All regressions include year fixed effects to control for time-varying changes in reporting practices and regulatory attention, and standard errors are clustered by issuer to account for within-firm correlation across observations. Firm-years for which financial data are unavailable are excluded from the regression sample, reducing the analytical sample from 3,416 to 2,373 firm-years. Variable definitions are reported in Appendix I.

4. Findings

4.1 S&E Content of Overlapping and Outlet-Specific Disclosures

Table 4.1 reports the topic structure of the three disclosure categories. The topic model identifies 21 topics in the overlapping corpus, 24 topics in the AIF-specific corpus, and 18 topics in the MD&A-specific corpus. The AIF-specific corpus contains the largest number of sentences (94,731) and exhibits the highest noise share (36.4%), suggesting greater thematic heterogeneity than either overlapping or MD&A-specific disclosures. By contrast, the overlapping corpus shows the greatest concentration of content, with the five largest topics accounting for 59.7% of all overlapping sentences. The corresponding coverage falls to 44.8% for AIF-specific disclosures and 33.0% for MD&A-specific disclosures, indicating that outlet-specific disclosures are distributed across a broader range of topics. While these topics do not exhaust the thematic content of each group, they capture the most systematically recurring S&E content and provide a basis for comparing the thematic composition of the three disclosure categories.

Table 4.1 Topic Structure of Three Disclosure Categories

	Overlapping	AIF-specific	MD&A-specific
Number of topics	21	24	18
Total sentences	27,610	94,731	5,396
Noise share (%)	23.1	36.4	24.9
Top 5 coverage (%)	59.7	44.8	33.0

Note: This table reports the topic structure for overlapping, AIF-specific, and MD&A-specific S&E disclosures. Topics are estimated separately for each disclosure category using BERTopic with FinBERT-ESG sentence embeddings. "Noise share" refers to the proportion of sentences assigned to the HDBSCAN noise cluster. "Top 5 topic coverage" reports the share of sentences assigned to the five largest non-noise topics in each disclosure category.

Table 4.2 reports the topic content, representative sentences, sentence counts, and proportions. The overlapping disclosures are concentrated in five dominant topics: Environmental Assessment Compliance, Community Proximity, Closure Plan Requirements, Carbon Emission

Management, and Workplace Safety as a Core Value. The environmental assessment topic repeatedly refers to environmental impact reports, assessment procedures, and approval processes, while the community proximity topic describes the nearest towns or communities and their distance from project sites. Similarly, the closure topic focuses on requirements to maintain closure plans and provide financial assurance, and the carbon emissions topic discusses regulatory limits, compliance obligations, and future emissions costs. The workplace safety topic frames health and safety through company policy, legislative compliance, and the language of "core values" and "working culture." Read together, these topics focus on the formal conditions under which mining projects are assessed, located, permitted, regulated, and operated.

Table 4.2 S&E Content of Three Disclosure Categories

Panel A. Overlapping Disclosures			
Topic Label	Representative Sentences	N	%
Environmental Assessment Compliance	Mining related activities in Argentina require the submission and approval of environmental impact report. Federal and provincial government decisions on proceeding with the project will be made following completion of the environmental assessment process.	6,536	23.7
Community Proximity	The nearest town to the project is Asunción Mita, a community of about [X] people, approximately [X] kilometers from the project. The closest all-season accessible community to the McFaulds Lake project area is Nakina, [X] km to the south.	5,828	21.1
Closure Plan Requirements	The mine permit requires that closure plans be updated every five years, and an accumulating closure fund is in place. We are required by various governments to provide financial assurance sufficient to allow a third party to implement approved closure and reclamation plans.	1,917	6.9
Carbon Emission Management	The regulations will create a greenhouse gas emissions limit for the province and a market-based approach to facilitating greenhouse gas reductions and regulatory compliance. The climate plan places some certainty on the future greenhouse gas (ghg) costs, while the limit on oil sands emissions anticipates that companies will be forced to ensure only the most profitable and efficient projects are developed.	1,204	4.4
Workplace Safety as a Core Value	It is the policy of the company to maintain safe working conditions at all its work sites, comply with health and safety legislation, and ensure that all employees are trained and comply with safety procedures. Safe and healthy workplace health and safety are among our core values, and part of our working culture.	1,003	3.6
Panel B: AIF-Specific Disclosures			

Topic Label	Representative Sentences	N	%
Project-Level S&E Impact Assessment	In February, the company submitted a coordinated final environmental assessment/environmental impact study for the CT gold project in accordance with the requirements of both the province of Ontario and the Government of Canada. The project received conditional environmental impact assessment approval and subsequently a second EIA was submitted in respect of modifications of the project.	14,840	15.7
Community Relations Strategy	The company is continually engaging with affected individuals and groups on its operational plans and activities to ensure it maintains its social licence to operate. The company's communication strategy includes having representatives formally meet with the community and local stakeholders on a regular basis.	10,165	10.7
Water and Waste Management	Waste disposal areas have been identified and approved in the ovoot tolgoi mining licence. These tailings also appear to affect water quality in Kenogamisis Lake.	7,808	8.2
Emissions Reduction Targets	The company's near-term approach to GHG mitigation is to focus on the energy efficiency of operations, with a longer-term plan and targets for GHG reduction under development. The company published the results of an independent report on the operational carbon footprint of the project, estimated to rank as one of the lowest in Canada and the world.	6,475	6.8
Indigenous Land Rights	The district is located in an area which is subject to recognized aboriginal land claims. The updated resource is located on the muskowekwan first nations home reservation and on newly designated treaty land entitlement (tle) lands.	3,215	3.4

Panel C: MD&A-Specific Disclosures

Topic Label	Representative Sentences	N	%
Community Resettlement Progress	In addition, victory nickel entered into a memorandum of understanding with the waswanipi cree first nation (wcfn) whereby the parties have agreed to work together to support development of the lac rocher deposit in a way that respects the collective interests of victory nickel, the wcfn and other stakeholders. The community has agreed in principle to enter into a community agreement with the company and this agreement is expected to be ratified within a month.	416	7.7
Local Employment Outcomes	The project will create over [X] full-time jobs for the community. [X]% of site-based employees were Indigenous citizens employed in a variety of roles including geology labourer, assay lab technician, and heavy equipment operator.	402	7.4
Reclamation Liability Estimates	The company's provision for reclamation and closure cost obligations represents management's best estimate of the present value of the future cash outflows required to settle the liability. Reclamation and closure costs have been estimated based on the company's interpretation of current regulatory requirements; however, changes in regulatory requirements may result in revisions to estimates.	391	7.2

Water Quality Management Measures	We continue to implement the water quality management measures required by the Water Quality Plan. The objective of the assessment is to evaluate various solutions and develop budgets to find a means of financing, implementing and maintaining potable water standards."	329	6.1
Safety Performance Metrics	The corporation has a zero harm policy which is applied at all sites, and continuous efforts are made to reduce the lost time injury frequency rate (LTIFR) at all the operations. Over the last two years, Gear has had zero lost time injuries and zero recordable injuries.	250	4

Note: This table reports the five largest non-noise BERTopic topics within each disclosure category. Topic labels are assigned interpretively after reviewing representative sentences generated by the model. Sentence counts and percentages are calculated within each disclosure category. Representative sentences are provided to illustrate the substantive content of each topic and are not intended to exhaust the full range of sentences assigned to the topic. Numerical values and dates are replaced with [X] during preprocessing to reduce identifying information.

AIF-specific disclosures exhibit a different thematic profile. The largest topics concern Project-Level S&E Impact Assessment, Community Relations Strategy, Water and Waste Management, Emissions Reduction Targets, and Indigenous Land Rights. The representative sentences show that this content is more detailed and project-specific than the overlapping disclosures. For example, one sentence refers to the submission of a "coordinated final environmental assessment/environmental impact study" for a named project, while another describes regular meetings with "the community and local stakeholders" as part of a communication strategy. Water and waste disclosures similarly refer to approved waste disposal areas and water quality effects, while emissions disclosures discuss operational carbon footprint assessment and longer-term GHG reduction targets. The Indigenous land rights topic further locates projects within recognized Aboriginal land claims, First Nations reserve lands, and treaty land entitlement arrangements. Read together, these topics suggest that AIF-specific disclosures carry more detailed accounts of how mining projects are assessed, negotiated, managed, and situated within local environmental, stakeholder, and land-rights contexts.

MD&A-specific disclosures reveal a third pattern. The dominant topics include Community Resettlement Progress, Local Employment Outcomes, Reclamation Liability Estimates, Water Quality Management Measures, and Safety Performance Metrics. Unlike the previous two groups, the representative sentences are strongly oriented toward implementation, outcomes, and managerial assessments. For example, the resettlement topic reports progress in relocation activities, including cases where community relocation was "completed by the end of June" or where community agreements were expected to be ratified within a specified timeframe. The employment topic reports the number of jobs created or the proportion of Indigenous employees working in operational roles. The reclamation topic is framed through management estimates of future obligations, including the "present value of the future cash outflows" required to settle closure and reclamation liabilities. Water-related disclosures similarly emphasize the implementation of water quality management measures and the development of financing and maintenance plans for potable water systems. The safety performance topic reports both quantified safety metrics, such as lost time injury frequency rates and total recordable injury rates, and zero-harm objectives. Rather than describing project characteristics or broader stakeholder relationships, these disclosures primarily report actions undertaken, outcomes achieved, measures implemented, and obligations estimated during the reporting period.

Taken together, the findings suggest that the three disclosure categories occupy distinct positions within the reporting system. Overlapping disclosures primarily communicate the regulatory status, geographic positioning, and formal obligations associated with mining projects. AIF-specific disclosures focus more heavily on project governance, stakeholder relationships, environmental management practices, and Indigenous land-rights arrangements. MD&A-specific disclosures, in contrast, emphasize implementation progress, operational outcomes, and

management estimates. The analysis therefore provides an initial indication of how different kinds of S&E matters are distributed across connected reporting documents.

As a supplementary analysis, Appendix H reports a pooled BERTopic model estimated on all three disclosure categories combined. The pooled model identifies 68 topics with a noise share of 40.4%, reflecting the greater heterogeneity of the combined corpus. Overlapping sentences are most concentrated in risk disclosure (41.0%), environmental damage and liability (29.2%), and GHG emissions (23.1%) topics. AIF-specific sentences dominate waste area management (88.1%), species and habitat protection (87.7%), and closure plan approval (86.0%), reflecting situated technical and ecological content absent from the other two groups. MD&A-specific sentences are most prominent in community agreements (20.5%), reclamation costs (17.8%), and occupational health and safety (14.3%), consistent with the implementation- and outcome-oriented emphasis identified in the main analysis. Overall, the pooled analysis supports the main interpretation that the three disclosure categories are not distributed randomly across S&E content, but are associated with different reporting emphases within the AIF–MD&A reporting system.

4.2 Accountability Logics Underlying Overlapping and AIF-Specific Disclosures

This section shows how concepts are repeatedly connected within disclosure sentences and therefore provides insight into the accountability logics associated with different documentary locations.

Table 4.3 reports network statistics for the three disclosure categories. The AIF-specific network is substantially larger than the overlapping and MD&A-specific networks, containing more nodes and connections and exhibiting lower edge concentration. This pattern suggests that AIF-specific disclosures connect S&E matters to a broader range of concepts and associations,

whereas overlapping and MD&A-specific disclosures are organized around a more limited set of relationships.

Table 4.3 Network Statistics

	Overlapping	AIF-Specific	MD&A-Specific
Nodes	41	140	19
Edges	33	194	27
Total edge weight	1,753	6,008	302
Edge concentration	0.109	0.050	0.086

Note: This table reports descriptive statistics for the phrase co-occurrence networks constructed separately for overlapping, AIF-specific, and MD&A-specific disclosures. Nodes represent extracted phrases, and edges represent co-occurrence within the same disclosure sentence. Edge weights are based on the number of distinct firms in which a phrase pair co-occurs, rather than raw sentence counts. Network concentration captures the extent to which total edge weight is concentrated in the strongest phrase-pair connections.

Table 4.4 reports the five strongest phrase connections in each disclosure category. For overlapping disclosures, Panel A shows that the strongest associations repeatedly connect environmental and social matters to regulatory obligations, legal exposure, and formal assessment processes. Representative sentences note that environmental damage may result in "possible legal liability," that environmental legislation may lead to "increased fines and penalties for non-compliance," and that environmental protection issues are linked to land-use regulation and permitting requirements. Other disclosures connect community impacts to environmental studies and permitting processes and frame climate change through greenhouse gas emissions regulation. Taken together, these phrase associations suggest that overlapping disclosures frequently render S&E matters accountable through compliance requirements, regulatory oversight, and legal consequences.

Table 4.4 Top Phrase Connections in Three Disclosure Categories

Panel A. Top phrase Connections in Overlapping S&E Content

Phrase Pair	Representative Sentence
environmental damage — legal liability	Such occurrences could result in environmental damage to properties of the company or others, delays, monetary losses and possible legal liability.

environmental protection — land use	Government regulations including regulations relating to land use, environmental protection, and mine permitting are subject to change.
environmental legislation — fines and penalties	Environmental legislation is evolving in a manner that may require stricter standards and enforcement, increased fines and penalties for non-compliance.
community impact — environmental studies	Environmental studies, permitting and social or community impact: the environmental impact statement was previously submitted to provincial and federal regulators.
climate change — greenhouse gas emissions	The corporation acknowledges climate change and that the increased regulation of greenhouse gas emissions may adversely affect operations.

Panel B. Top Phrase Connections in AIF-Specific S&E Content

Phrase Pair	Representative Sentence
air quality — hazardous waste	These conditions include programs for the recovery and relocation of flora, reforestation, recovery and relocation of fauna, monitoring of surface water quality, monitoring of air quality and hazardous waste management.
indigenous peoples — united nations	The company entered into impact benefit agreements with the matachewan first nation as the young davidson mine is situated within the traditional territory of these two first nations.
health and safety — human rights	Pursuant to the modern slavery act, entities are required to file public reports annually on steps taken to prevent and reduce the use of forced labour and child labour in their supply chains.
first nation — traditional territory	The company operates in some areas currently or traditionally inhabited or used by indigenous peoples and subject to indigenous rights or claims.
child labour — forced labour	The federal and provincial governments have a duty to consult with aboriginal people on actions and decisions that may affect the asserted aboriginal or treaty rights.

Panel C. Top Phrase Connections in MD&A-Specific S&E Content

Phrase Pair	Representative Sentence
reclamation costs — site closure	Mining interests represent capitalized expenditures related to the development of mining properties, acquisition costs, and estimated site closure and reclamation costs.
health and safety — capital expenditures	Zapata mitigates these business risks by: having assets in several diverse fields; maintaining cost effective operations; maintaining a balance between oil and gas properties,; operating our own properties to control the amount and timing of capital expenditures; using new technology to maximize production and recoveries and reduce operating costs; ensuring staff are trained and in compliance with current regulatory and health and safety requirements; and drilling wells in areas with multiple high deliverability zone potential.
asset retirement obligations — reclamation costs	The provision for asset retirement obligations represents management's best estimate of the present value of the future abandonment and reclamation costs required under current regulatory requirements.

first nation — closure plan	Treasury has entered into long term relationship agreements with the eagle lake first nation and wabauskang first nation to ensure effective communication during the environmental assessment process.
frequency rate — total recordable injury	Our rate of total recordable injury frequency rate increased in compared to the prior year; we have conducted an analysis of all injuries to identify trends and areas for improvement at each mine.

Note: This table reports the five strongest phrase-pair connections in each disclosure category. Edge weight is measured as the number of distinct firms in which the phrase pair co-occurs within the same sentence. Representative sentences are selected from sentences containing the reported phrase pair and are used to ground the interpretation of the associated accountability logic.

For AIF-specific disclosures, Panel B points to a broader rights- and responsibility-oriented set of associations. The strongest phrase connections link S&E matters to Indigenous peoples, traditional territory, human rights, forced and child labour, hazardous waste, and air quality. The representative sentences ground these connections in impact benefit agreements with First Nations, Indigenous rights and claims, modern slavery reporting requirements, and environmental exposure. Rather than emphasizing compliance and legal consequences alone, AIF-specific disclosures make S&E matters accountable through relationships, rights, labour standards, and environmental responsibilities embedded within particular social and institutional contexts.

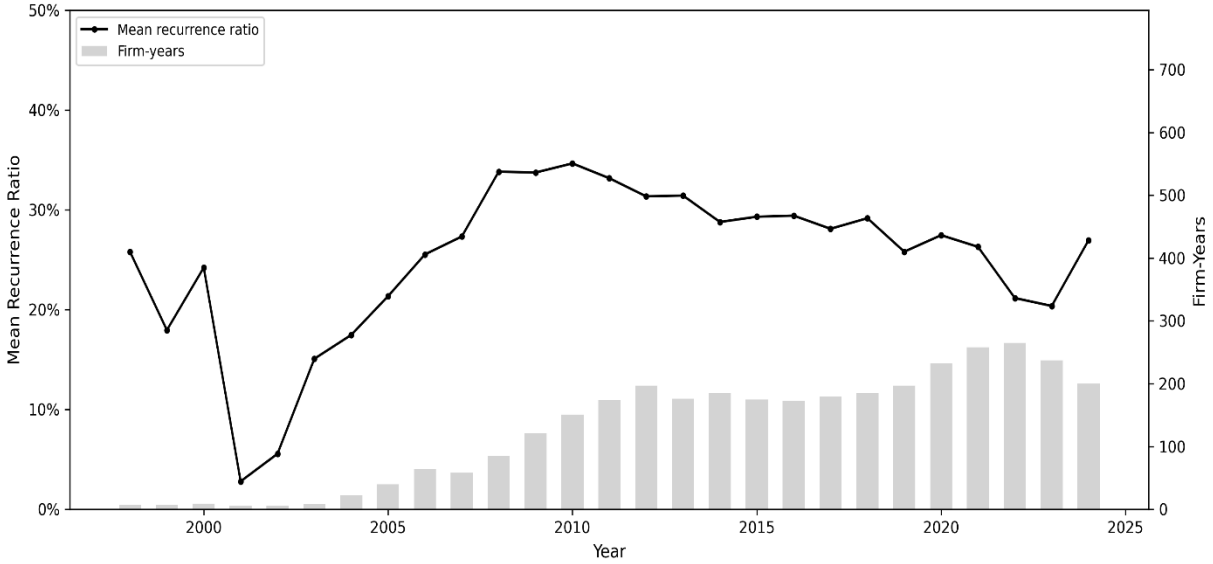
MD&A-specific disclosures show a different pattern again. In Panel C, the strongest associations center on reclamation costs, site closure, asset retirement obligations, health and safety requirements, capital expenditures, and injury frequency metrics. Representative sentences refer to estimated site closure and reclamation costs, the "present value" of future abandonment and reclamation obligations, compliance with health and safety requirements, and changes in total recordable injury frequency rates. These associations suggest that MD&A-specific disclosures render S&E matters accountable through operational performance, managerial estimation, and measurable outcomes.

Taken together, the network results suggest three distinct accountability logics. For overlapping disclosures, the dominant logic is regulatory-compliance: S&E matters are made answerable through requirements, approvals, penalties, and legal exposure. In AIF-specific disclosures, the logic shifts toward rights and responsibility, with S&E matters connected to Indigenous rights, stakeholder relationships, labour standards, human rights, and environmental responsibilities. MD&A-specific disclosures rely more on a performance-and-estimation logic, where answerability is grounded in implementation activities, operational metrics, expenditures, and estimates of future obligations. This helps explain why recurrence is selective: the S&E accounts that recur across AIF and MD&A filings are especially associated with regulatory and compliance-based forms of account-giving, while rights-based and performance-based accounts more often remain tied to a specific reporting outlet.

4.3 Temporal and Firm-Level Variation in Cross-Outlet Recurrence

This section examines whether the extent of AIF–MD&A S&E disclosure recurrence varies systematically across firms and over time. Figure 4.1 presents the annual average recurrence ratio together with the number of firm-year observations in each year. The figure shows substantial fluctuation during the early years of the sample, when the number of observations is relatively small. As sample coverage expands after the mid-2000s, the recurrence ratio rises and then stabilizes at a higher level, generally fluctuating between approximately 25 and 35 percent. The pattern suggests that cross-outlet recurrence became more common over time.

Figure 4.1 Average Recurrence Ratio over Time



Note: This figure reports the annual mean recurrence ratio together with the number of firm-year observations for each fiscal year. The recurrence ratio is calculated as the number of overlapping S&E sentences divided by the total number of AIF S&E sentences in the corresponding firm-year. Higher values indicate that a larger proportion of AIF S&E disclosure recurs in the corresponding MD&A filing. The observation count is reported to provide context for year-to-year variation in the recurrence ratio.

To provide a context for the cross-sectional analysis, Table 4.5 reports descriptive statistics for the filing-level recurrence ratio and the firm characteristics used in the regression analysis. The mean recurrence ratio is 0.289, indicating that, on average, approximately 29% of S&E sentences in an AIF filing recur in substantively similar form in the corresponding MD&A filing. The ratio also exhibits substantial variation across firm-years, ranging from 0.000 to 1.000, with an interquartile range from 0.111 to 0.414. The explanatory variables similarly show meaningful cross-sectional variation. Firm size (LOG_ASSETS) ranges from 0.210 to 9.555, while leverage ranges from 0.000 to 1.299. Profitability (ROA) includes both profitable and loss-making firms, ranging from -3.244 to 0.443. PPE, CAPEX, filing length, and MD&A S&E disclosure volume also vary substantially across observations, indicating heterogeneity in firms' operating structure, investment activity, and disclosure characteristics.

Table 4.5 Descriptive Statistics

Variable	N	Mean	SD	Min	P25	Median	P75	Max
RECURRENCE_RATIO	2,373	0.289	0.223	0.000	0.111	0.229	0.414	1.000
LOG_ASSETS	2,373	5.098	1.946	0.210	3.680	4.900	6.556	9.555
ROA	2,373	-0.164	0.461	-3.244	-0.174	-0.052	0.015	0.443
LEVERAGE	2,373	0.121	0.219	0.000	0.000	0.013	0.164	1.299
TURNOVER	2,373	0.200	0.305	0.000	0.000	0.001	0.312	1.615
PPE	2,373	0.790	0.601	0.000	0.535	0.762	0.938	4.665
CAPEX	2,373	0.145	0.210	0.000	0.029	0.083	0.168	1.416
MTB	2,373	2.284	3.604	-1.719	0.756	1.284	2.374	25.951
LOG_PAGES	2,373	4.242	0.437	3.178	3.932	4.234	4.554	5.328
LOG_MDA_SE_COUNT	2,373	2.001	0.992	0.000	1.386	2.079	2.708	3.970

Note: This table reports descriptive statistics for the recurrence ratio and firm-level variables used in the regression analysis. The recurrence ratio is calculated as the number of overlapping AIF sentences divided by the total number of AIF S&E sentences in the filing. Financial variables are measured at the firm-year level. Variable definitions are provided in Appendix I.

Table 4.6 reports the fractional logistic regression results. Because the model is estimated using a fractional logit specification, the coefficients indicate how each variable is associated with the log-odds of the conditional mean of the recurrence ratio, rather than a direct percentage-point change. Firm size (LOG_ASSETS) is negatively associated with recurrence (coefficient = -0.183, $p < 0.01$), indicating that, holding other variables constant, larger firms tend to have a lower expected recurrence ratio. Filing length shows an even stronger negative association (coefficient = -0.797, $p < 0.01$), suggesting that longer AIF filings contain proportionally more S&E content that remains specific to the AIF rather than recurring in the MD&A. PPE is also negatively associated with recurrence (coefficient = -0.084, $p < 0.10$), providing evidence that more asset-intensive firms have lower recurrence ratios.[27]

[27] Appendix J reports OLS estimates as a robustness check. The results are broadly consistent with the fractional logistic specification. In particular, firm size and filing length remain negatively associated with recurrence, while MD&A S&E disclosure volume remains positively associated with recurrence.

Table 4.6 Fractional Logistic Regression Results

	RECURRENCE_RATIO
LOG_ASSETS	−0.183*** (0.029)
ROA	0.117* (0.065)
LEVERAGE	0.136 (0.144)
TURNOVER	−0.071 (0.104)
PPE	−0.084* (0.044)
CAPEX	−0.181 (0.125)
MTB	−0.000 (0.007)
LOG_PAGES	−0.797*** (0.110)
LOG_MDA_SE_COUNT	1.030*** (0.045)
Constant	2.345*** (0.713)
Year FE	Yes
Clustered SE (firm)	Yes
N	2,373

*Note: This table reports fractional logistic regression estimates for the association between firm characteristics and the filing-level recurrence ratio. The dependent variable is the recurrence ratio, bounded between zero and one. All specifications include year fixed effects. Standard errors are clustered by issuer. Coefficients are reported in log-odds units. Statistical significance is indicated as follows: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.*

By contrast, LOG_MDA_SE_COUNT is strongly and positively associated with recurrence (coefficient = 1.030, $p < 0.01$). This indicates that firms with more extensive S&E disclosure in the MD&A tend to exhibit higher levels of cross-outlet recurrence. ROA is also positively associated with recurrence (coefficient = 0.117, $p < 0.10$). Leverage, turnover, capital expenditures, and market-to-book ratio are not statistically significant, suggesting that recurrence

is not systematically associated with firms' financing structure, operating turnover, investment intensity, or market valuation after controlling for the other variables.

Taken together, the RQ3 results show that recurrence is neither constant across time nor randomly distributed across firms. The temporal analysis suggests that recurrence became more prevalent as the reporting system matured, although outlet-specific disclosures remained substantial throughout the sample period. The regression results further indicate that recurrence is associated with characteristics of disclosure scope and reporting structure. The positive association with MD&A S&E volume indicates that greater S&E disclosure volume in the MD&A is associated with higher cross-outlet recurrence. By contrast, the negative association with firm size is consistent with larger firms having a wider range of AIF-specific content, potentially reflecting greater diversity in projects, jurisdictions, stakeholder relationships, and environmental obligations. The negative association with filing length points in the same direction: longer AIFs appear to contain more differentiated S&E information that does not recur in the MD&A. The negative association with PPE may similarly reflect the more site-specific operational and remediation issues faced by asset-intensive firms, which are less likely to recur across both reporting outlets. Taken together, these associations suggest that greater disclosure breadth and operational complexity may correspond to greater outlet-specificity rather than greater cross-outlet recurrence.

5. Discussion

This study examines how S&E disclosure is organized across connected corporate reporting documents in the Canadian mining sector. Taken together, the findings show that cross-outlet recurrence is highly selective. The S&E accounts that recur across both AIF and MD&A filings differ systematically from those that remain outlet-specific, both in their substantive content

and in the forms of account-giving through which they are rendered answerable. Overlapping disclosures are primarily associated with regulatory-compliance accounts, whereas AIF-specific and MD&A-specific disclosures are more strongly associated with rights-and-responsibility and performance-and-estimation accounts, respectively. The analysis further shows that recurrence became more common over time but remained far from complete, and that firms differ substantially in the extent to which S&E accounts recur across reporting outlets. Rather than representing a neutral duplication of information, recurrence appears concentrated in particular forms of S&E account within the AIF–MD&A reporting system.

The study makes three theoretical contributions. First, it contributes to accounting research on inscription by showing that making an issue reportable is not the endpoint of inscription. Prior accounting research has emphasized how inscriptions transform dispersed activities, events, and relations into forms that can be compared, combined, and acted upon (Latour, 1987; Robson, 1992; Qu and Cooper, 2011). This study extends that perspective by showing that reportable S&E matters acquire different degrees of documentary durability within corporate reporting. Some accounts become recurrent across related filings, while others remain tied to a particular reporting outlet. The contribution is therefore to shift attention from whether S&E matters are inscribed into corporate reports to how far those inscriptions become durable within a reporting system.

Second, the study contributes to accountability research by examining how different forms of answerability are organized within a connected reporting system. Prior accountability research has shown that different accountability relationships invoke different obligations, expectations, and grounds for justification, and may coexist in tension with one another (Roberts, 1991; Sinclair, 1995; Ahrens, 1996). Accountability has therefore been examined as a multifaceted phenomenon that is experienced through different relationships, responsibilities, and styles of account-giving.

This study extends that literature by shifting attention from accountability forms themselves to their documentary organization. The findings suggest that different forms of answerability are not distributed evenly across corporate reporting. Regulatory-compliance, rights-and-responsibility, and performance-and-estimation logics exhibit different patterns of recurrence and outlet-specificity within the AIF–MD&A reporting system.

This perspective also provides a new way of thinking about the limits of accountability. Prior SEA research has often located those limits in the selective content of disclosures, the exclusion of stakeholder voices, or the institutional constraints embedded within reporting practices (Gray et al., 1997; O'Dwyer, 2005; Cooper and Owen, 2007; Messner, 2009). The findings here suggest an additional consideration. Different forms of account-giving do not acquire the same degree of documentary presence within the reporting system. The contribution is therefore to show that the limits of accountability may arise not only from what organizations disclose, but also from how different forms of answerability are organized across corporate reporting.

Third, the study contributes to SEA disclosure research by introducing a documentary perspective on the distinction between environmental and social disclosure. Prior studies have distinguished environmental and social disclosures in terms of their content, determinants, legitimacy pressures, and informational effects (Cormier and Gordon, 2001; Gray et al., 2001; Deegan et al., 2002; Cormier et al., 2011), treating them as different categories of reported information. This study identifies an additional dimension through which the distinction can be understood: their position within the documentary architecture of corporate reporting. The topic analysis reveals that several environmental matters are disproportionately represented among overlapping disclosures that recur across both the AIF and MD&A, while social matters are

disproportionately concentrated in outlet-specific disclosures that remain tied to a single filing genre. The study therefore shows that the environmental-social distinction has a documentary dimension: the two categories of disclosure occupy different positions within corporate reporting architecture and are subject to different conditions of visibility, circulation, and accountability.

This study has limitations that suggest future research opportunities. First, its focus on Canadian mining firms and AIF–MD&A filings provides a useful setting for examining cross-outlet S&E disclosure, but future studies could test whether similar patterns appear in other industries, jurisdictions, or reporting regimes. Second, the sentence-level design captures repeated disclosure statements, but it may miss correspondence expressed through tables, quantitative measures, or broader narrative structures. Third, because the internal preparation process is not observed, the recurrence of S&E content across AIF and MD&A filings may be associated with several mechanisms: deliberate disclosure strategy, shared internal templates, or external counsel boilerplate. The present design cannot distinguish between these explanations, and future research examining the production process through interviews or archival evidence would help clarify the mechanisms underlying cross-outlet recurrence. Finally, linking cross-outlet disclosure patterns to environmental incidents, enforcement actions, or community conflicts would help evaluate how documentary mobility relates to substantive accountability.

Ultimately, this study suggests that S&E disclosure cannot be understood only by examining what firms disclose within individual reports. It also matters how disclosures are distributed across the reporting architecture. Corporate reporting does not simply make S&E matters visible; it shapes the forms in which they become recurrent, outlet-specific, and available for comparison across connected documents. The critical question, then, is not only whether S&E matters enter corporate reports, but what forms they take once they are translated into reporting

practice. Understanding this process is essential for grasping how corporate reporting both extends and limits the visibility of S&E matters.

Chapter 5: Conclusion

This dissertation began from a simple premise: corporate disclosures are not neutral transmissions of information. They are representational practices through which organizations make particular activities, relationships, and outcomes visible while leaving others in the background (Burchell et al., 1980; Hopwood, 1987). A substantial body of accounting research has shown that disclosure is shaped by legitimacy pressures, institutional expectations, and impression management incentives (Patten, 1991; Neu et al., 1998; Deegan et al., 2002; Cho and Patten, 2007; Nazari et al., 2015). That literature has generated important insights into why firms disclose, how much they disclose, and how disclosure characteristics such as tone and readability shape stakeholder perceptions and market outcomes. At the same time, it has been less attentive to a related set of questions about how disclosure narratives are constructed in practice: what firms choose to foreground within their accounts, how particular narratives are sustained or allowed to disappear over time, and how related accounts are distributed across the connected documents that together constitute a reporting system. This dissertation addressed these questions by examining strategic narratives in regulated disclosures through three linked dimensions: content, temporality, and coordination.

Across the three studies, several findings emerge. The first study shows that CD&A disclosures vary systematically in their substantive emphasis. The analysis identifies 19 topics across eight thematic categories, with accounting performance, aligning managerial and shareholder interests, and pay-for-performance among the most prominent. Changes in accounting performance are also associated with differences in topical emphasis: increases in ROA are associated with greater emphasis on aligning interests and compliance and less emphasis on

accounting, pay-for-performance, peers, independence, and risk management. Additional analyses identify further variation in topic choices associated with declining performance, changes in CEO pay, and earnings management. These findings are consistent with substantive topic selection representing a content-based dimension of impression management alongside the stylistic characteristics more commonly examined in disclosure research.

The second study shows that voluntary social disclosure also varies in how content develops across reporting periods. Introduced disclosure tends to concern new developments and organizational responses, carried-forward disclosure centers more on established systems and commitments, and abandoned disclosure tends to be more episodic or time-bound. At the filing level, disclosure change also varies with topic composition, with greater emphasis on Labour Relations and Workplace Health and Safety associated with higher overall rates of introduction and abandonment. Greater institutional alignment is associated with greater disclosure continuity overall, although the strength of this relationship varies with the relative prominence of selected social topics. These findings distinguish the substantive character of content occupying different lifecycle stages from the broader disclosure dynamics associated with the composition of the reporting portfolio.

The third study shows that S&E accounts are distributed selectively across connected reporting outlets. Overlapping, AIF-specific, and MD&A-specific disclosures differ in both substantive content and accountability logics. Overlapping disclosure is associated primarily with formalized environmental and regulatory matters and a regulatory-compliance logic; AIF-specific disclosure places greater emphasis on situated environmental and stakeholder concerns and reflects a rights-and-responsibility logic; and MD&A-specific disclosure focuses more on implementation, operational outcomes, and managerial estimates and reflects a performance-and-estimation logic.

Cross-outlet recurrence has also increased over time and varies systematically with filing and firm characteristics. Together, these findings indicate that the content of corporate accounts differs not only across firms and reporting periods, but also across their positions within the reporting system.

Taken together, the three studies make three broader contributions to accounting research. The first contribution of the dissertation is to develop a more comprehensive understanding of narrative visibility in corporate disclosure. Prior research has examined disclosure through measures such as quantity, quality, tone, readability, and market informativeness, showing that corporate reports are shaped not only by informational demands but also by managerial efforts to influence how organizational performance is understood (Aerts, 2005; Li, 2008; Huang et al., 2014; Nazari et al., 2017). This dissertation extends that literature by showing that disclosure strategy also involves choices about what becomes visible within the narrative itself. Across the three studies, these choices appear in the selective emphasis of compensation topics, the continuation or abandonment of social narratives, and the repeated or selective appearance of social and environmental accounts across reporting outlets.

The findings further extend narrative visibility along temporal and documentary dimensions. Prior research on social and environmental reporting has documented changes in disclosure volume and content over time, often in relation to legitimacy threats and stakeholder demands (Gray et al., 1995; Tilling and Tilt, 2010). However, reports are still often treated as annual snapshots, and disclosure change is commonly measured as an increase or decrease in the amount of information disclosed. The lifecycle analysis shows that particular disclosure content can instead be introduced, carried forward, or abandoned across successive reporting periods. This distinction directs attention to the durability of individual narratives rather than only to aggregate changes in disclosure from one year to the next.

Similarly, much accounting research examines one report at a time, whether an annual report, a proxy statement, a sustainability report, or an MD&A filing (Laksmana et al., 2012; Boiral, 2013; Kent and Zunker, 2013; Brown et al., 2024). This approach has produced important insights, but it can obscure the fact that firms communicate through a set of connected reporting documents that serve different accountability purposes and address different audiences. The cross-document analysis shows that related accounts may recur across reporting outlets or remain specific to particular documents. Taken together, the three studies therefore extend narrative visibility beyond the presence or prominence of information within an individual report. Visibility also concerns the substantive issues receiving emphasis, the durability of particular narratives over time, and the distribution of related accounts across the wider reporting system.

The second contribution concerns corporate accountability. The preceding distinction is important because differences in narrative visibility correspond to differences in the accounts made available for external scrutiny. The cross-document analysis provides the clearest evidence of this distinction: overlapping disclosure is associated with a regulatory-compliance logic, whereas AIF-specific and MD&A-specific disclosures are associated with rights-and-responsibility and performance-and-estimation logics, respectively. Thus, even within the same broad domain of S&E disclosure, different forms of account-giving are associated with different positions within the reporting system. The other two studies extend this perspective by showing that firms also vary in the substantive issues receiving emphasis and in whether particular narratives remain part of the reporting portfolio over time. Collectively, these findings suggest that accountability through disclosure involves not only the provision of information, but also the organization of the accounts through which organizational activities, relationships, and responsibilities become available for scrutiny.

This perspective also complicates a simple association between greater disclosure and greater accountability. A larger quantity of disclosure may still concentrate attention on a limited range of issues, while repeated disclosure may maintain an established account without necessarily providing new information. Conversely, content that appears only temporarily or within a particular reporting outlet may make visible matters that receive less attention elsewhere. Disclosure quantity, temporal continuity, and cross-document recurrence therefore capture different characteristics of account-giving and should not be treated as interchangeable indications of stronger or weaker accountability. Attention to how corporate narratives are organized consequently provides a more differentiated basis for examining the possibilities and limitations of disclosure as a form of accountability.

The third contribution is methodological. The dissertation shows how large-scale computerized textual analysis can be used not as a substitute for interpretation, but as a way of identifying patterned variation within large blocks of disclosure text. Across the three studies, topic modeling, sentence embeddings, similarity measures, phrase co-occurrence networks, and related large-data techniques make it possible to identify variation in thematic emphasis, trace the continuity and abandonment of disclosure content, and examine the recurrence of related accounts across reporting outlets at a scale that would be difficult to achieve through manual reading alone. The methodological contribution of this dissertation therefore lies in combining computational identification of textual variation with interpretive analysis of representative sentences, topic meanings, lifecycle patterns, and accountability logics. In doing so, the dissertation uses computational methods to open large disclosure corpora to theoretically informed interpretation rather than treating them as self-sufficient sources of meaning.

As ESG reporting expands in scope, ambition, and scrutiny, the questions raised in this dissertation will become more pressing. Firms are being asked to account for more issues, across more documents, to more audiences, and over longer time horizons. This makes it important to understand how disclosures are constructed: what firms foreground, what they sustain, and where they position particular accounts. Making these choices visible is part of the broader effort to understand what corporate accountability can and cannot do. This dissertation has sought to make a start.

References

Abeysekera, I. (2014). Why do firms disclose and not disclose structural intangibles? *Advances in Accounting*, *30*(2), 381-393.

Aerts, W. (1994). On the use of accounting logic as an explanatory category in narrative accounting disclosures. *Accounting, Organizations and Society*, *19*(4-5), 337-353.

Aerts, W. (2005). Picking up the pieces: Impression management in the retrospective attributional framing of accounting outcomes. *Accounting, Organizations and Society, 30*(6), 493-517.

Aerts, W., & Yan, B. (2017). Rhetorical impression management in the letter to shareholders and institutional setting: A metadiscourse perspective. *Accounting, Auditing & Accountability Journal*, *30*(2), 404-432.

Ahrens, T. (1996). Styles of accountability. *Accounting, Organizations and Society*, *21*(2-3), 139-173.

Alias, M. S., Fuad, M. H., Hoong, X. L. F., & Hin, E. G. Y. (2023). Financial text categorisation with FinBERT on key audit matters. *2023 IEEE Symposium on Computers & Informatics*, 63-69.

Baier, P., Berninger, M., & Kiesel, F. (2020). Environmental, social and governance reporting in annual reports: A textual analysis. *Financial Markets, Institutions & Instruments*, *29*(3), 93-118.

Baker, A. C., Larcker, D. F., McClure, C. G., Saraph, D., & Watts, E. M. (2024). Diversity washing. *Journal of Accounting Research*, *62*(5), 1661-1709.

Balsam, S., Boone, J., Liu, H., & Yin, J. (2016). The impact of say-on-pay on executive compensation. *Journal of Accounting and Public Policy, 35*(2), 162-191.

Bao, Y., & Datta, A. (2014). Simultaneously discovering and quantifying risk types from textual risk disclosures. *Management Science, 60*(6), 1371-1391.

Basu, S., Ma, X., & Briscoe-Tran, H. (2022). Measuring multidimensional investment opportunity sets with 10-K text. *The Accounting Review, 97*(1), 51-73.

Beattie, V. (2014). Accounting narratives and the narrative turn in accounting research: Issues, theory, methodology, methods and a research framework. *The British Accounting Review*, *46*(2), 111-134.

Bellstam, G., Bhagat, S., & Cookson, J. A. (2021). A text-based analysis of corporate innovation. *Management Science*, *67*(7), 4004-4031.

Bellucci, M., Acuti, D., Simoni, L., & Manetti, G. (2021). Restoring an eroded legitimacy: the adaptation of nonfinancial disclosure after a scandal and the risk of hypocrisy. *Accounting, Auditing & Accountability Journal*, *34*(9), 195-217.

Bianchi, P. A., Causholli, M., Minutti-Meza, M., & Sulcaj, V. (2023). Social networks analysis in accounting and finance. *Contemporary Accounting Research*, *40*(1), 577-623.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research, 3*(Jan), 993-1022.

Bloomfield, M. J. (2021). Compensation disclosures and strategic commitment: Evidence from revenue-based pay. *Journal of Financial Economics*, *141*(2), 620-643.

Bochkay, K., Brown, S. V., Leone, A. J., & Tucker, J. W. (2023). Textual analysis in accounting: What's next? *Contemporary Accounting Research*, *40*(2), 765-805.

Boiral, O. (2013). Sustainability reports as simulacra? A counter-account of A and A+ GRI reports. *Accounting, Auditing & Accountability Journal*, *26*(7), 1036-1071.

Bozzolan, S., Cho, C. H., & Michelon, G. (2015). Impression management and organizational audiences: The Fiat group case. *Journal of Business Ethics*, *126*, 143-165.

Brown, S. V., Hinson, L. A., & Tucker, J. W. (2024). Financial statement adequacy and firms' MD&A disclosures. *Contemporary Accounting Research*, *41*(1), 126-162.

Brown, S. V., & Knechel, W. R. (2016). Auditor–client compatibility and audit firm selection. *Journal of Accounting Research*, *54*(3), 725-775.

Brown, S. V., & Tucker, J. W. (2011). Large-sample evidence on firms' year-over-year MD&A modifications. *Journal of Accounting Research*, *49*(2), 309-346.

Burchell, S., Clubb, C., Hopwood, A., Hughes, J., & Nahapiet, J. (1980). The roles of accounting in organizations and society. *Accounting, Organizations and Society*, *5*(1), 5-27.

Cadman, B., Carter, M. E., & Peng, X. (2021). The participation constraint and CEO equity grants. *The Accounting Review, 96*(1), 67-89.

Campbell, B. (2012). Corporate Social Responsibility and development in Africa: Redefining the roles and responsibilities of public and private actors in the mining sector. *Resources Policy*, *37*(2), 138-143.

Campello, R. J., Moulavi, D., & Sander, J. (2013). Density-based clustering based on hierarchical density estimates. *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 160-172.

Cannon, J. N., Ling, Z., Wang, Q., & Watanabe, O. V. (2020). 10-K disclosure of corporate social responsibility and firms' competitive advantages. *European Accounting Review*, *29*(1), 85-113.

Carley, K. (1993). Coding choices for textual analysis: A comparison of content analysis and map analysis. *Sociological Methodology*, 75-126.

Carmona, S., Ezzamel, M., & Gutiérrez, F. (2002). The relationship between accounting and spatial practices in the factory. *Accounting, Organizations and Society*, *27*(3), 239-274.

Cha, Y., Gill, S., & Wong-On-Wing, B. (2023). Clawback policy enforcement: To disclose or not to disclose. *Advances in Accounting*, *62*, Article 100661.

Chelli, M., Durocher, S., & Richard, J. (2014). France's new economic regulations: insights from institutional legitimacy theory. *Accounting, Auditing & Accountability Journal*, *27*(2), 283-316.

Chen, J. C., & Roberts, R. W. (2010). Toward a more coherent understanding of the organization–society relationship: A theoretical consideration for social and environmental accounting research. *Journal of Business Ethics*, *97*(4), 651-665.

Cho, C. H., Michelon, G., & Patten, D. M. (2012). Impression management in sustainability reports: An empirical investigation of the use of graphs. *Accounting and the Public Interest*, *12*(1), 16-37.

Cho, C. H., Michelon, G., Patten, D. M., & Roberts, R. W. (2015). CSR disclosure: the more things change…? *Accounting, Auditing & Accountability Journal*, *28*(1), 14-35.

Cho, C. H., & Patten, D. M. (2007). The role of environmental disclosures as tools of legitimacy: A research note. *Accounting, Organizations and Society*, *32*(7-8), 639-647.

Cho, C. H., Roberts, R. W., & Patten, D. M. (2010). The language of US corporate environmental disclosure. *Accounting, Organizations and Society*, *35*(4), 431-443.

Cooper, S. M., & Owen, D. L. (2007). Corporate social reporting and stakeholder accountability: The missing link. *Accounting, Organizations and Society*, *32*(7-8), 649-667.

Cooper, S., & Slack, R. (2015). Reporting practice, impression management and company performance: A longitudinal and comparative analysis of water leakage disclosure. *Accounting and Business Research*, *45*(6-7), 801-840.

Cormier, D., & Gordon, I. M. (2001). An examination of social and environmental reporting strategies. *Accounting, Auditing & Accountability Journal*, *14*(5), 587-617.

Cormier, D., Ledoux, M. J., & Magnan, M. (2011). The informational contribution of social and environmental disclosures for investors. *Management Decision*, *49*(8), 1276-1304.

Cowen, S. S., Ferreri, L. B., & Parker, L. D. (1987). The impact of corporate characteristics on social responsibility disclosure: A typology and frequency-based analysis. *Accounting, Organizations and Society*, *12*(2), 111-122.

Deegan, C., & Rankin, M. (1996). Do Australian companies report environmental news objectively? An analysis of environmental disclosures by firms prosecuted successfully by the Environmental Protection Authority. *Accounting, Auditing & Accountability Journal*, *9*(2), 50-67.

Deegan, C., & Rankin, M. (1997). The materiality of environmental information to users of annual reports. *Accounting, Auditing & Accountability Journal*, *10*(4), 562-583.

Deegan, C., Rankin, M., & Tobin, J. (2002). An examination of the corporate social and environmental disclosures of BHP from 1983-1997: A test of legitimacy theory. *Accounting, Auditing & Accountability Journal*, *15*(3), 312-343.

De Franco, G., Fogel-Yaari, H., & Li, H. (2020). MD&A textual similarity and auditors. *Auditing: A Journal of Practice & Theory*, *39*(3), 105-131.

Demers, E., Wang, V. X., & Wu, K. (2024). Measuring corporate human capital disclosures: Lexicon, data, code, and research opportunities. *Journal of Information Systems*, *38*(2), 163-186.

Diaz-Rainey, I., Griffin, P. A., Lont, D. H., Mateo-Márquez, A. J., & Zamora-Ramírez, C. (2024). Shareholder Activism on Climate Change: Evolution, Determinants, and Consequences: I. Diaz-Rainey et al. *Journal of Business Ethics*, *193*(3), 481-510.

DiMaggio, P., Nag, M., & Blei, D. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of US government arts funding. *Poetics*, *41*(6), 570-606.

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, *48*(2), 147-160.

Doidge, C., Karolyi, G. A., & Stulz, R. M. (2007). Why do countries matter so much for corporate governance? *Journal of Financial Economics*, *86*(1), 1-39.

Dyer, T., Lang, M., & Stice-Lawrence, L. (2017). The evolution of 10-K textual disclosure: Evidence from latent Dirichlet allocation. *Journal of Accounting and Economics, 64*(2-3), 221-245.

Elshandidy, T., & Kamel, H. (2024). Tone of narrative disclosures and earnings management: UK evidence. *Advances in Accounting*, *64*, Article 100710.

Ezzamel, M., & Robson, K. (1995). Accounting in time: organizational time-reckoning and accounting practice. *Critical Perspectives on Accounting*, *6*(2), 149-170.

Farshadfar, S., Samarbakhsh, L., & Jahan, I. (2025). Executive age, executive gender and financial statement comparability. *Advances in Accounting*, *68*, Article 100805.

Frankel, R., Mayew, W. J., & Sun, Y. (2010). Do pennies matter? Investor relations consequences of small negative earnings surprises. *Review of Accounting Studies*, *15*(1), 220-242.

Gipper, B. (2021). The economic effects of expanded compensation disclosures. *Journal of Accounting and Economics, 71*(1), 101338.

Global Reporting Initiative Standards. (2016). Standards. *Retrieved from https://www.globalreporting.org/how-to-use-the-gri-standards/gri-standards-english-language/*

Godfrey, J., Mather, P., & Ramsay, A. (2003). Earnings and impression management in financial reports: The case of CEO changes. *Abacus, 39*(1), 95-123.

Gray, R. (2007). Taking a long view on what we now know about social and environmental accountability and reporting. *Issues in Social and Environmental Accounting*, *1*(2), 169-198.

Gray, R., Dey, C., Owen, D., Evans, R., & Zadek, S. (1997). Struggling with the praxis of social accounting: Stakeholders, accountability, audits and procedures. *Accounting, Auditing & Accountability Journal*, *10*(3), 325-364.

Gray, R., Kouhy, R., & Lavers, S. (1995). Corporate social and environmental reporting: a review of the literature and a longitudinal study of UK disclosure. *Accounting, Auditing & Accountability Journal*, *8*(2), 47-77.

Gray, R., Javad, M., Power, D. M., & Sinclair, C. D. (2001). Social and environmental disclosure and corporate characteristics: A research note and extension. *Journal of Business Finance & Accounting*, *28*(3-4), 327-356.

Grootendorst, M. (2022), "BERTopic: Neural topic modeling with a class-based TF-IDF procedure", *arXiv:2203.05794*.

Hastie, T., Qian, J., & Tay, K. (2023). *An introduction to Glmnet. Retrieved from https://cloud.r-project.org/web/packages/glmnet/vignettes/glmnet.pdf*

Hemmings, D., Hodgkinson, L., & Williams, G. (2020). It's OK to pay well, if you write well: The effects of remuneration disclosure readability. *Journal of Business Finance & Accounting*, *47*(5-6), 547-586.

Herremans, I. M., Nazari, J. A., & Mahmoudian, F. (2016). Stakeholder relationships, engagement, and sustainability reporting. *Journal of Business Ethics*, *138*(3), 417-435.

Hoberg, G., & Lewis, C. (2017). Do fraudulent firms produce abnormal disclosure? *Journal of Corporate Finance, 43,* 58-85.

Hooghiemstra, R., Kuang, Y. F., & Qin, B. (2017). Does obfuscating excessive CEO pay work? The influence of remuneration report readability on say-on-pay votes. *Accounting and Business Research, 47*(6), 695-729.

Hopwood, A. G. (1987). The archeology of accounting systems. *Accounting, Organizations and Society*, *12*(3), 207-234.

Huang, A. H., Wang, H., & Yang, Y. (2023). FinBERT: A large language model for extracting information from financial text. *Contemporary Accounting Research*, *40*(2), 806-841.

Huang, X., Teoh, S. H., & Zhang, Y. (2014). Tone management. *The Accounting Review*, *89*(3), 1083-1113.

Jenkins, H., & Yakovleva, N. (2006). Corporate social responsibility in the mining industry: Exploring trends in social and environmental disclosure. *Journal of Cleaner Production*, *14*(3-4), 271-284.

Kent, P., & Zunker, T. (2013). Attaining legitimacy by employee information in annual reports. *Accounting, Auditing & Accountability Journal*, *26*(7), 1072-1106.

Kim, H., Lee, E., & Yoo, D. (2023). Do SEC filings indicate any trends? Evidence from the sentiment distribution of forms 10-K and 10-Q with FinBERT. *Data Technologies and Applications*, *57*(2), 293-312.

Kim, J. B., Wang, C., & Wu, F. (2023). The real effects of risk disclosures: Evidence from climate change reporting in 10-Ks. *Review of Accounting Studies*, *28*(4), 2271-2318.

Kim, N. K. W., Kwon, S., & Shin, J. Y. (2025). Do compensation committees do what they say they do? Evidence from narrative disclosure in proxy statements. *Finance Research Letters*, *73*, Article 106637.

Klein Jr, V. H., & Reilley, J. T. (2024). The temporal dynamics of enterprise risk management. *Critical Perspectives on Accounting*, *99*, 102363.

Kuang, Y. F., & Qin, B. (2009). Performance-vested stock options and interest alignment. *The British Accounting Review*, *41*(1), 46-61.

Laksmana, I. (2008). Corporate board governance and voluntary disclosure of executive compensation practices. *Contemporary Accounting Research*, *25*(4), 1147-1182.

Laksmana, I., Tietz, W., & Yang, Y. W. (2012). Compensation discussion and analysis (CD&A): Readability and management obfuscation. *Journal of Accounting and Public Policy, 31*(2), 185-203.

Lang, M., & Stice-Lawrence, L. (2015). Textual analysis and international financial reporting: Large sample evidence. *Journal of Accounting and Economics*, *60*(2-3), 110-135.

Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Harvard University Press.

Lauwo, S. G., Otusanya, O. J., & Bakre, O. (2016). Corporate social responsibility reporting in the mining sector of Tanzania: (Lack of) government regulatory controls and NGO activism. *Accounting, Auditing & Accountability Journal*, *29*(6), 1038-1074.

LaViers, L., Sandvik, J., & Xu, D. (2024). CEO pay ratio voluntary disclosures and stakeholder reactions. *Review of Accounting Studies*, *29*(1), 109-150.

Leung, S., Parker, L., & Courtis, J. (2015). Impression management through minimal narrative disclosure in annual reports. *British Accounting Review*, *47*(3), 275-289.

Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics, 45*(2-3), 221-247.

Li, X., & Yang, H. I. (2016). Mandatory financial reporting and voluntary disclosure: The effect of mandatory IFRS adoption on management forecasts. *The Accounting Review*, *91*(3), 933-953.

Liang, C., Lourie, B., Nekrasov, A., & Yoo, I. S. (2025). Voluntary disclosure of workforce gender diversity. *Journal of Financial Reporting*, *10*(1), 101-125.

Lin, Y., Shen, R., Wang, J., & Julia Yu, Y. (2024). Global evolution of environmental and social disclosure in annual reports. *Journal of Accounting Research*, *62*(5), 1941-1988.

Liu, M., Taylor-Neu, K., Saxton, G. D., Neu, D., Rahaman, A. S., & Everett, J. (2025). Indigenous peoples, environmental accountability and the semantic meaning of resource extraction firm disclosures. *Accounting, Auditing & Accountability Journal*, *38*(5), 1375-1404.

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, *66*(1), 35-65.

Loughran, T., McDonald, B., & Otteson, J. R. (2023). How have corporate codes of ethics responded to an era of increased scrutiny? *Journal of Business Ethics*, *183*(4), 1029-1044.

Lowe, A. et al. (2020). Problematizing profit and profitability: discussions. *Accounting, Auditing & Accountability Journal*, *33*(4), 753-793.

Luft Mobus, J. (2005). Mandatory environmental disclosures in a legitimacy theory context. *Accounting, Auditing & Accountability Journal*, *18*(4), 492-517.

Martin, X., Xu, J. J., & Zhou, G. (2021). Does compensation matter? Evidence from CD&A disclosures. *Retrieved from https://ssrn.com/abstract=3819394*

Matsumoto, D., Pronk, M., & Roelofsen, E. (2011). What makes conference calls useful? The information content of managers' presentations and analysts' discussion sessions. *The Accounting Review*, *86*(4), 1383-1414.

Nazari, J. A., Herremans, I. M., & Warsame, H. A. (2015). Sustainability reporting: external motivators and internal facilitators. *Corporate Governance*, *15*(3), 375-390.

Nazari, J. A., Hrazdil, K., & Mahmoudian, F. (2017). Assessing social and environmental performance through narrative complexity in CSR reports. *Journal of Contemporary Accounting & Economics*, *13*(2), 166-178.

McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv:1802.03426*.

Merkley, K. J. (2014). Narrative disclosure and earnings performance: Evidence from R&D disclosures. *The Accounting Review*, *89*(2), 725-757.

Messner, M. (2009). The limits of accountability. *Accounting, Organizations and Society*, *34*(8), 918-938.

Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology*, *83*(2), 340-363.

Michelon, G., Pilonato, S., & Ricceri, F. (2015). CSR reporting practices and the quality of disclosure: An empirical analysis. *Critical Perspectives on Accounting*, *33*, 59-78.

Michelon, G., Rodrigue, M., & Trevisan, E. (2020). The marketization of a social movement: Activists, shareholders and CSR disclosure. *Accounting, Organizations and Society*, *80*, Article 101074.

Miller, P. (1998). The margins of accounting. *European Accounting Review*, *7*(4), 605-621.

Miller, P., & O'leary, T. (1994). Accounting, "economic citizenship" and the spatial reordering of manufacture. *Accounting, Organizations and Society*, *19*(1), 15-43.

Miller, P., & Rose, N. (1990). Governing economic life. *Economy and society*, *19*(1), 1-31.

Min, W., Liang, W., Yin, H., Wang, Z., Li, M., & Lal, A. (2021). Explainable deep behavioral sequence clustering for transaction fraud detection. *arXiv:2101.04285*.

Moreno, A., Jones, M. J., & Quinn, M. (2019). A longitudinal study of the textual characteristics in the chairman's statements of Guinness: An impression management perspective. *Accounting, Auditing & Accountability Journal*, *32*(6), 1714-1741.

Morgan-Lopez, A., Kim, A., Chew, R., & Ruddle, P. (2017). Predicting age groups of Twitter users based on language and metadata features. *PLOS ONE, 12*(8), e0183537.

Mutti, D., Yakovleva, N., Vazquez-Brust, D., & Di Marco, M. H. (2012). Corporate social responsibility in the mining industry: Perspectives from stakeholder groups in Argentina. *Resources Policy*, *37*(2), 212-222.

Neu, D. (2000). "Presents" for the "Indians": Land, colonialism and accounting in Canada. *Accounting, Organizations and Society*, *25*(2), 163-184.

Neu, D., & Graham, C. (2004). Accounting and the holocausts of modernity. *Accounting, Auditing & Accountability Journal*, *17*(4), 578-603.

Neu, D., Liu, M., & Saxton, G. D. (2026). Not monstrous but procedural: Accounting for cruelty to First Nations children. *Accounting, Auditing & Accountability Journal*, *pre-print*.

Neu, D., & Saxton, G. D. (2022). Building ethical narratives: The audiences for AICPA editorials. *Journal of Business Ethics, 182*(4), 1055-1072.

Neu, D., & Saxton, G. D. (2023). Twitter bots, democratic deliberation and social accountability: The case of #OccupyWallStreet. *Accounting, Auditing & Accountability Journal*, *38*(3), 934-964.

Neu, D., Saxton, G. D., Everett, J., & Rahaman, A. (2022). The centrality of ethical utterances within professional narratives. *Accounting History, 27*(1), 75-94.

Neu, D., Warsame, H., & Pedwell, K. (1998). Managing public impressions: Environmental disclosures in annual reports. *Accounting, Organizations and Society, 23*(3), 265-282.

O'Dwyer, B. (2005). The construction of a social account: a case study in an overseas aid agency. *Accounting, Organizations and Society*, *30*(3), 279-296.

Oliver, C. (1991). Strategic responses to institutional processes. *Academy of Management Review*, *16*(1), 145-179.

Ontario Securities Commission. (2015). *51-102 Continuous Disclosure Obligations*. *Retrieved from https://osc.ca/en/securities-law/instruments-rules-policies/5/51-102.*

Papke, L. E., & Wooldridge, J. M. (1996). Econometric methods for fractional response variables with an application to 401 (k) plan participation rates. *Journal of Applied Econometrics*, *11*(6), 619-632.

Patten, D. M. (1991). Exposure, legitimacy, and social disclosure. *Journal of Accounting and Public Policy*, *10*(4), 297-308.

Patten, D. M. (1992). Intra-industry environmental disclosures in response to the Alaskan oil spill: A note on legitimacy theory. *Accounting, Organizations and Society*, *17*(5), 471-475.

Patten, D. M. (2002). The relation between environmental performance and environmental disclosure: a research note. *Accounting, Organizations and Society*, *27*(8), 763-773.

Packard, H. A., Pawliczek, A., & Skinner, A. N. (2023). Voluntary performance disclosures in the CD&A. *The Accounting Review*, *98*(6), 435-466.

Parker, L. D. (2004). 'Presenting' the past: perspectives on time for accounting and management history. *Accounting, Business & Financial History*, *14*(1), 1-27.

Peterson, K., Schmardebeck, R., & Wilks, T. J. (2015). The earnings quality and information processing effects of accounting consistency. *The Accounting Review*, *90*(6), 2483-2514.

Preston, A. M., Wright, C., & Young, J. J. (1996). Imag [in] ing annual reports. *Accounting, Organizations and Society*, *21*(1), 113-137.

Preston, A. M., Chua, W. F., & Neu, D. (1997). The diagnosis-related group-prospective payment system and the problem of the government of rationing health care to the elderly. *Accounting, Organizations and Society*, *22*(2), 147-164.

Qiu, J. X., Faulkner, A., & Can, A. E. (2021). Towards theme detection in personal finance questions. *arXiv:2110.01550*.

Qu, S. Q., & Cooper, D. J. (2011). The role of inscriptions in producing a balanced scorecard. *Accounting, Organizations and Society*, *36*(6), 344-362.

Rao, D., Yarowsky, D., Shreevats, A., & Gupta, M. (2010). Classifying latent user attributes in Twitter. *International Conference on Information and Knowledge Management,* 37–44.

Richardson, V. J., & Waegelein, J. F. (2002). The influence of long-term performance plans on earnings management and firm performance. *Review of Quantitative Finance and Accounting*, *18*, 161-183.

Roberts, J. (1991). The possibilities of accountability. *Accounting, Organizations and Society*, *16*(4), 355-368.

Roberts, J., & Scapens, R. (1985). Accounting systems and systems of accountability—understanding accounting practices in their organisational contexts. *Accounting, Organizations and Society*, *10*(4), 443-456.

Roberts, R. W. (1992). Determinants of corporate social responsibility disclosure: An application of stakeholder theory. *Accounting, Organizations and Society*, *17*(6), 595-612.

Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). Stm: An R package for structural topic models. *Journal of Statistical Software, 91,* 1-40.

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science, 58*(4), 1064-1082.

Robson, K. (1992). Accounting numbers as "inscription": Action at a distance and the development of accounting. *Accounting, Organizations and Society*, *17*(7), 685-708.

Safari, M., & Parker, L. D. (2024). Understanding multiple accountability logics within corporate governance policy discourse: Resistance, compromise, or selective coupling? *European Accounting Review*, *33*(4), 1467-1496.

Schlenker, B. R., Forsyth, D. R., Leary, M. R., & Miller, R. S. (1980). Self-presentational analysis of the effects of incentives on attitude change following counterattitudinal behavior. *Journal of Personality and Social Psychology*, *39*(4), 553.

Securities and Exchange Commission. (2006). *Executive compensation and related person disclosure*. Release Nos. 33-8732A; 34-54302*A.*

Sinclair, A. (1995). The chameleon of accountability: Forms and discourses. *Accounting, Organizations and Society*, *20*(2-3), 219-237.

Soobaroyen, T., & Mahadeo, J. D. (2016). Community disclosures in a developing country: insights from a neo-pluralist perspective. *Accounting, Auditing & Accountability Journal*, *29*(3), 452-482.

Thorne, L., S. Mahoney, L., & Manetti, G. (2014). Motivations for issuing standalone CSR reports: A survey of Canadian firms. *Accounting, Auditing & Accountability Journal*, *27*(4), 686-714.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *58*(1), 267-288.

Tilling, M. V., & Tilt, C. A. (2010). The edge of legitimacy: Voluntary social and environmental reporting in Rothmans' 1956-1999 annual reports. *Accounting, Auditing & Accountability Journal*, *23*(1), 55-81.

van Bommel, K. (2014). Towards a legitimate compromise? An exploration of integrated reporting in the Netherlands. *Accounting, Auditing & Accountability Journal*, *27*(7), 1157-1189.

Vo, T. T. N., & Canil, J. M. (2019). CEO pay disparity: Efficient contracting or managerial power? *Journal of Corporate Finance*, *54*, 168-190.

Wakefield, R. (2008). Networks of accounting research: A citation-based structural and network analysis. *The British Accounting Review*, *40*(3), 228-244.

Wang, I. Y., Wang, X., & Wangerin, D. (2020). Consequences of increased compensation disclosure transparency: Evidence from CEO pay in acquiring firms. *Journal of Accounting, Auditing & Finance, 35*(4), 667-695.

Yates, J., & Orlikowski, W. J. (1992). Genres of organizational communication: A structurational approach to studying communication and media. *Academy of Management Review*, *17*(2), 299-326.

Yates, J., & Orlikowski, W. (2002). Genre systems: Structuring interaction through communicative norms. *The Journal of Business Communication*, *39*(1), 13-35.

Appendices

Appendix A. Variable Definitions

Variables	Definitions
Independent variables	
Main analyses	
$TOPIC_i$	An indicator variable that equals one if at least one FREX word of topic i appears in a sentence, and zero otherwise. This definition applies to both individual topics and broader thematic groupings. Examples of individual topics include long-term incentives, bonus calculation, and market survey sources. Examples of combined thematic categories include aligning interests, pay-for-performance, and peers.
Additional analyses	
$TOPIC_PROP_i$	A continuous variable measured as the total number of FREX words for topic i in the entire CD&A section, scaled by the total word count in that CD&A. This definition applies to both individual topics and broader thematic groupings. The variable is tested in Section 6.3.
TOPIC_ABNi	An indicator variable which equals one if the overall FREX word count of topic i in a given CD&A filing exceeds its average FREX word count in that firm's prior filings by 20%, and zero otherwise. This definition applies to both individual topics and broader thematic groupings. The variable is tested in Section 6.3.
Dependent variables	
Main analyses	
ROA_INCREASE	An indicator variable that equals one if return on assets for the fiscal year increases from the previous year, and zero otherwise. Return on assets is calculated as net income divided by beginning-of-year total assets.
Additional analyses	
CEOPAY_INCREASE	An indicator variable that equals one if the total CEO compensation increases in the current year, and zero otherwise. The variable is tested in Section 6.5.
JMB	An indicator variable that equals one if a firm's earnings just meet or slightly exceed analysts' consensus earnings forecast by 2 cents or less, and zero otherwise. The variable is tested in Section 6.6.
CAR	A continuous variable measured by firms' three-day cumulative abnormal returns around their CD&A filing dates. The variable is tested in Section 6.7.

ROA_INCREASE_LEAD	An indicator variable that equals one if return on assets for the following fiscal year increases relative to the current year, and zero otherwise. The variable is tested in Section 6.8.
Control variables	
ASSETS	The natural logarithm of total assets at the beginning of the fiscal year.
LEVERAGE	Total debt (short- and long-term) divided by beginning-of-year total assets.
TURNOVER	The natural logarithm of the ratio of total revenue over the fiscal year to beginning-of-year total assets.
MTB	The ratio of market value to book value of total assets, as of the beginning of the fiscal year. Market value of total assets is calculated as the sum of market value of equity, book value of total liabilities, and liquidation value of preferred stock minus deferred tax liabilities.
PPE	Net property, plant, and equipment divided by beginning-of-year total assets.
CAPEX	Capital expenditures divided by beginning-of-year total assets.
EXECPAY	The ratio of total executive compensation to total revenue over the year, multiplied by 100.
LENGTH	Total word count of the CD&A divided by 1000.

Note: Appendix A presents definitions for the variables in the main and additional analyses.

Appendix B. Social Keyword Dictionary and Exclusion Criteria

This appendix presents the social keyword dictionary (Panel A) and the exclusion criteria (Panel B) used to identify voluntary social sentences in MIC filings. Panel A lists all keywords and phrases used to extract candidate social sentences, organised by thematic category. Panel B presents the exclusion criteria (governance and boilerplate blacklist) applied after keyword matching. These patterns identify sentences that contain social keywords but appear in non-social contexts — primarily compensation plan descriptions, director biography tables, legal disclaimers, and standardized governance language. Patterns use case-insensitive regular expression matching. Mandatory board gender diversity disclosures (final row) are excluded to preserve the voluntary-only nature of the sample.

Panel A. Social Keyword Dictionary

Category	Keywords and Phrases
Diversity & Inclusion	board diversity; gender diversity; diversity and inclusion; diversity & inclusion; diversity program; diversity initiative; diversity goal; diversity target; diversity policy; diversity strategy; diversity committee; diversity in the workplace; diversity in our; diversity of our; diversity metrics; diversity data; diversity reporting; diversity training; diversity hiring; diversity recruitment; diversity efforts; promote diversity; advance diversity; committed to diversity; foster diversity; value diversity; support diversity; embrace diversity; champion diversity; encourage diversity; diverse and inclusive; diverse workforce; diverse board; diverse team; diverse talent; diverse candidates; commitment to diversity; importance of diversity; benefits of diversity; culture of diversity; gender parity; pay equity; pay gap; gender gap; gender-based; gender based; women in leadership; women on board; women on the board; female director; female directors; women directors; women executives; underrepresented group; underrepresented groups; equal opportunity employer; employment equity plan; employment equity program; nondiscrimination policy; anti-discrimination; diversity disclosure; nondiscrimination; eeo; lgbtq; lgbt; transgender
Labor Standards	collective bargaining; living wage; minimum wage; equal pay; fair wage; occupational health and safety; occupational health & safety; health and safety program; health and safety policy; health and safety management; workplace safety program; workplace safety policy; worker safety; sexual harassment policy; anti-harassment policy; harassment policy; harassment-free; forced labor; forced labour; modern slavery; modern slavery act; child labor; child labour; fair labor standards; fair labour standards; labor rights; labour rights; worker rights; employee rights; ilo convention; ilo standard; ilo guideline; health and safety; workplace harassment; zero tolerance; lost time injury; lost-time injury; injury rate; injury frequency; safety record; safety performance; ilo; eicc; fla; ruggie; absenteeism; harassment; labour; labor; workforce

Employee Wellbeing	employee wellbeing; employee wellness; employee well-being; employee mental health; work-life balance; work life balance; employee engagement program; employee engagement survey; employee satisfaction survey; employee retention program; employee retention strategy; parental leave policy; parental leave program; maternity leave; paternity leave; childcare benefit; childcare program; child care benefit; employee assistance program; employee development program; employee development plan; professional development program; learning and development; talent development program; employee training program
Community	community investment program; community investment fund; social investment; community engagement program; community engagement initiative; community development program; community development fund; indigenous community; indigenous communities; indigenous peoples; indigenous people; first nations community; first nations communities; local community investment; host community; volunteer program; employee volunteer; community volunteer; charitable donation; charitable contribution; charitable giving; corporate giving; community giving; social responsibility program; corporate social responsibility program; community benefit; community benefits agreement; community partnership; nonprofit organization; non-profit organization; philanthropic initiative; philanthropic program; corporate philanthropy; poverty reduction; philanthropy; philanthropic; corporate donation; corporate donations; donation match; donation program; indigenous; donation; donations; community; communities
Human Rights & CSR	human rights policy; human rights program; human rights due diligence; human rights assessment; human rights commitment; human rights report; human rights risk; supply chain responsibility; responsible sourcing; supplier code of conduct; supplier standards; modern slavery statement; corporate social responsibility; vulnerable populations; vulnerable communities; veterans program; veterans initiative; disability inclusion; disability program; accessibility program; accessibility plan; accommodation policy; csr; human rights

Panel B: Exclusion Criteria

Exclusion Category	Patterns	Rationale
Executive compensation boilerplate	named executive officer; compensation committee; stock option plan; annual base salary; base salary; Black-Scholes; option pricing model; grant-date fair value; optionee ceases to be; permanently disabled or dying; optionee retiring; Earned ($); Share-based awards; Option-based awards; Non-equity incentive; Pension Value	Sentences describing executive compensation structures; triggered by social keywords (e.g., health, disability) but

		unrelated to voluntary social disclosure
Clinical and medical context	safety and tolerability; safety and efficacy; safety and pharmacokinetic; clinical trial; clinical study; oral administration; intravenous	Pharmaceutical and clinical research language; safety keywords appear in non-social disclosure context
COVID-19 boilerplate	COVID-19 pandemic; public health recommendation; contagious disease outbreak; precautionary measures in relation to the meeting; workforce productivity…pandemic/virus/covid; mitigate potential risk to the health and safety associated with	Pandemic-era virtual meeting protocols and risk disclosures; not substantive voluntary social content
Financial statement content	Current Assets; Current Liabilities; Working Capital; Accrued Interest; As at December	Balance sheet and financial table content erroneously triggered by social keywords
Governance committee descriptions	annual directors and officers questionnaire; among other things, responsible; responsible for overseeing various aspects; Audit Committee Agenda; as is typical for; appointed to serve as a member of; be it resolved; options granted under the plan; Governance and Administrative Matters; demonstrations of extraordinary personal commitment…community; monitor review and provide oversight…community; reviewing managements implementation…community; ensure that the company has in place…community; governance is not a static issue…community; specific responsibilities of the board…community	Director evaluation criteria and committee mandate boilerplate; community/safety keywords appear as incidental references to board responsibilities
Equity and financial terminology	equity compensation; equity offering; equity financing; equity securities; equity interest; equity stake; equity method; private equity; return on equity; shareholders equity; total equity; common equity; equity award; equity grant; equity plan; equity incentive	Financial equity terms; triggered by diversity-related keywords (e.g., equity, pay equity) but used

		in financial rather than social disclosure contexts
Labour/workforce noise	subject to laws/regulations…labour/labor; risks and hazards…labour/labor; cost of labour/labor; direct labour/labor; labour/labor disruption; relatively small workforce; small workforce; laboratory	Labour keywords appearing in operational risk disclosures, cost reporting, or unrelated scientific contexts (e.g., laboratory)
Community noise — governance and biography	investment community; analyst community; financial community; business community; shareholder community; scientific community; Health, Safety, Environment and Technical Committee; except as otherwise indicated…community; at the request of the board…community; directors of several non-profit; board member for several non-profit	Community keyword appearing in financial community references, director biography descriptions, or committee mandate listings rather than community relations disclosure
Community noise — COVID and legal	COVID/coronavirus/pandemic…community; Securities Commission…Box/Street/Avenue/Suite; right to dissent a shareholder; dissenter means a shareholder	Community keyword in pandemic risk context or legal regulatory address listings; not voluntary social disclosure
Director biography tables	public company directorships; common shares security ownership; fiscal year na common shares; while providing general accounting and tax services; member of the environmental health and safety committee; member of the health and safety committee	Director biography and compensation table content; social keywords appear as committee

		memberships or professional background descriptions rather than substantive disclosure
Mandatory board gender diversity disclosures	representation of women; number of women on the board; targets for the representation; has not adopted…target; identification and nomination of women; criteria for selecting…diversity; considers its size…diversity; Nomination of Directors…diversity; proportion of women on the board; proportion of women…director; women on the board…percent/proportion/number/director; board…include/includes/will include…woman/women; identified as…visible minority/woman/women/indigenous/disability; self-identify…designated group/visible minority/woman/women; designated group…woman/women/visible minority/indigenous/disability; diversity policy…board; board…diversity policy; written…diversity policy; adopted…diversity policy; gender diversity…board/director/candidate/nominee; board diversity; diversity…board of directors; board of directors…diversity; one/two/three/[number]…woman/women…board/director; proportion of…women/female…board/executive/officer	Sentences fulfilling mandatory disclosure requirements under Items 11–15 of Form 58-101F1 (effective December 31, 2014); excluded to preserve the voluntary-only nature of the sample

Appendix C. Robustness to Cosine Similarity Threshold

This appendix reports the sensitivity of the lifecycle classification to the choice of cosine similarity threshold τ. For each sentence in year t, the threshold determines whether the sentence is classified as INTRODUCED (maximum cosine similarity to any year-t−1 sentence below τ) or CARRIED_FORWARD (above τ). Symmetrically, for each sentence in year t−1, the threshold determines whether the sentence is classified as ABANDONED (maximum cosine similarity to any year-t sentence below τ). The primary specification uses τ = 0.80 (bold row). Results are reported for $\tau \in \{0.75, 0.80, 0.85\}$. The total sample size is identical across thresholds (N = 80,890 for the year *t* sample and N = 70,483 for the year *t*−1 sample). The main qualitative findings that higher-salience sentences are more likely to be carried forward and less likely to be abandoned are directionally consistent across all three thresholds.

Lifecycle Classification Rates by Cosine Similarity Threshold

Threshold (τ)	N (year t)	INTRODUCED (%)	CARRIED_FORWARD (%)	N (year t−1)	ABANDONED (%)
0.75	80,890	10.7%	89.3%	70,483	7.4%
0.80 [primary]	**80,890**	**15.9%**	**84.1%**	**70,483**	**11.1%**
0.85	80,890	24.0%	76.0%	70,483	17.1%

Appendix D. Institutional Salience Reference Corpus

This appendix lists the 43 reference sentences used to compute INST_ALIGNMENT. For each voluntary social sentence in the sample, INST_ALIGNMENT is defined as the maximum cosine similarity between the sentence's FinBERT-ESG embedding and the embeddings of these 43 reference sentences.

No.	Source	Reference Text
1	TMX/TSX ESG Primer (2020)	renewed attention to social factors including human capital management employee health and safety labour practices and human rights
2	TMX/TSX ESG Primer (2020)	effective management of social factors can produce benefits including an advantage in attracting and retaining talent
3	TMX/TSX ESG Primer (2020)	investors and other capital market participants are seeking disclosure of financial and non-financial information
4	TMX/TSX ESG Primer (2020)	increasing investor expectations for disclosure on how the board provides oversight of social factors
5	TMX/TSX ESG Primer (2020)	social licence to operate in local communities is a material factor for issuers
6	TMX/TSX ESG Primer (2020)	companies with strong social strategies focused on business performance will be better positioned to create long-term value
7	UN Global Compact Principles 1–6	businesses should support and respect the protection of internationally proclaimed human rights
8	UN Global Compact Principles 1–6	businesses should make sure that they are not complicit in human rights abuses
9	UN Global Compact Principles 1–6	businesses should uphold the freedom of association and the effective recognition of the right to collective bargaining
10	UN Global Compact Principles 1–6	businesses should uphold the elimination of all forms of forced and compulsory labour
11	UN Global Compact Principles 1–6	businesses should uphold the effective abolition of child labour
12	UN Global Compact Principles 1–6	businesses should uphold the elimination of discrimination in respect of employment and occupation
13	UN Global Compact Principles 1–6	respecting labour rights and advancing decent work means treating workers with dignity
14	UN Global Compact Principles 1–6	safe working conditions living wages freedom from discrimination and access to remedy
15	UN Global Compact Principles 1–6	human rights due diligence to identify prevent and mitigate adverse human rights impacts

16	UN Global Compact Principles 1–6	corporate sustainability starts with a company value system and a principles-based approach
17	GRI 400 Social Standards	new employee hires and employee turnover by age group gender and region
18	GRI 400 Social Standards	benefits provided to full-time employees including health care parental leave retirement provision
19	GRI 400 Social Standards	parental leave policies and return to work and retention rates of employees that took parental leave
20	GRI 400 Social Standards	occupational health and safety management system including worker participation consultation and communication
21	GRI 400 Social Standards	hazard identification risk assessment and incident investigation processes for occupational health and safety
22	GRI 400 Social Standards	work-related injuries ill health and fatalities including rates and root cause analysis
23	GRI 400 Social Standards	occupational health services provided to workers including health promotion programs
24	GRI 400 Social Standards	worker training on occupational health and safety topics
25	GRI 400 Social Standards	average hours of training per year per employee by gender and employee category
26	GRI 400 Social Standards	programs for upgrading employee skills and transition assistance programs
27	GRI 400 Social Standards	percentage of employees receiving regular performance and career development reviews
28	GRI 400 Social Standards	percentage of individuals within governance bodies and employees by gender age group and other indicators
29	GRI 400 Social Standards	ratio of basic salary and remuneration of women to men by employee category
30	GRI 400 Social Standards	incidents of discrimination and corrective actions taken
31	GRI 400 Social Standards	operations and suppliers in which the right to freedom of association and collective bargaining may be at risk
32	GRI 400 Social Standards	operations and suppliers considered to have significant risk for incidents of forced or compulsory labour
33	GRI 400 Social Standards	operations and suppliers considered to have significant risk for incidents of child labour
34	GRI 400 Social Standards	incidents of violations involving rights of indigenous peoples and actions taken

35	GRI 400 Social Standards	operations with local community engagement impact assessments and development programs
36	GRI 400 Social Standards	operations with significant actual and potential negative impacts on local communities
37	GRI 400 Social Standards	new suppliers that were screened using social criteria
38	GRI 400 Social Standards	significant actual and potential negative social impacts in the supply chain and actions taken
39	GRI 400 Social Standards	percentage of significant product and service categories for which health and safety impacts are assessed
40	TRC Calls to Action, s. 92	adopt the United Nations Declaration on the Rights of Indigenous Peoples as a reconciliation framework
41	TRC Calls to Action, s. 92	ensure that Aboriginal peoples have equitable access to jobs training and education opportunities
42	TRC Calls to Action, s. 92	provide education to management and staff on the history of Aboriginal peoples including the history of residential schools
43	TRC Calls to Action, s. 92	build respectful relationships with Indigenous peoples through meaningful consultation on all matters that affect them

Appendix E. Variable Definitions

Variable	Definition
Dependent Variable	
PCT_INTRODUCED	Proportion of sentences in the year t firm-year sample that are classified as INTRODUCED (maximum cosine similarity to any year t−1 sentence below $\tau = 0.80$). Bounded [0, 1].
PCT_ABANDONED	Proportion of sentences in the year t−1 firm-year sample that are classified as ABANDONED (maximum cosine similarity to any year-t sentence below $\tau = 0.80$). Bounded [0, 1].
Independent Variable	
INST_ALIGNMENT	Mean sentence-level institutional salience score for the firm-year. Sentence-level score is the maximum cosine similarity between a sentence's FinBERT-ESG embedding and the embeddings of the 43-sentence reference corpus (Appendix D). Higher values indicate greater semantic alignment with voluntary institutional frameworks. Computed separately for the year t and year t−1 estimation samples.
Controls	
$TOPIC_i$	Topic share calculated as the proportion of voluntary social sentences assigned to a given topic within a filing.
ROA	Net income divided by beginning-of-year total assets.
LEVERAGE	Total debt divided by beginning-of-year total assets.
MTB	Market value of equity plus book value of debt divided by book value of total assets.
PPE	Net property, plant, and equipment divided by beginning-of-year total assets.
LOG_ASSETS	Natural logarithm of total assets.
TURNOVER	Total revenue divided by beginning-of-year total assets.
CAPEX	Capital expenditure divided by beginning-of-year total assets.
AUDITOR_BIG4	Indicator equal to 1 if the firm is audited by one of the Big Four accounting firms (Deloitte, PwC, KPMG, or Ernst & Young), and 0 otherwise.

Appendix F. Representative Sentences by Topic and Lifecycle Stage

This appendix presents representative sentences for each of the five dominant social topics across the three lifecycle stages. Sentences are selected by identifying the sentence whose FinBERT-ESG embedding is closest to the centroid of the relevant topic-stage combination. Numerical values and dates are replaced with [X] during preprocessing to reduce identifying information.

$T_{Community_Philanthropy}$: Community and Philanthropy

Introduced — Community Outreach Initiatives

Falco continued its community outreach program and held, on [X], its initial project presentation and session in Rouyn-Noranda, which was attended by approximately individuals.

The company and its predecessors have regularly sponsored joint training initiatives to enhance the skills of local community members and have been a sponsor of Outland Youth Employment Camps, which provides Indigenous youth with jobs, education, and training in the forestry and natural resources sectors.

Carried Forward — Corporate Giving Programs

The company has adopted a donations policy, which establishes guidelines for the administration of the company's corporate giving program and outlines the four key areas of priority for philanthropic activities, which include education, healthcare, community development, and arts and culture.

The community engagement policies and programs that the company currently has in place constitute a robust framework for managing social risk and for providing tangible and substantial socioeconomic benefits to the local communities in which we operate.

Abandoned — Project-specific Community Activities

The company maintained an absolute minimal amount of project activities, most of which related to community relations in the Cañariaco area and communications with the various levels of government in Peru.

During XX, the company continued to invest in infrastructure, education, and sustainable development projects in Didipio and neighbouring communities.

T_{Health_Safety}: Workplace Health and Safety

Introduced — Safety Performance Updates

Except for that tragic incident, total recordable incident frequency and other safety performance measures were on target for the year.

Osisko also improved its safety performance at Canadian Malartic and completed funding of the million Canadian Malartic closure cost guarantee in the third quarter, a first in the Québec mining industry.

Carried Forward — Safety Management System

Osisko established a health and safety program applicable to all employees of Osisko and to those of its contractors working on the Canadian Malartic site.

Health and safety regulation affects the merged group's activities.

Abandoned — Past Safety Achievements

The successful construction startup and strong well deliverability increased yearly production results and was completed with a perfect safety record.

Significant safety efforts have been made to improve safety performance during the remaining construction and subsequent commissioning and operating stages in Nunavut.

$T_{DEI_Initiatives}$: DEI Initiatives

Introduced — Diversity Recruitment Efforts

As part of its strategy to recruit and maintain a diversified organization, the corporation will promote diversity within its team by giving preference to qualified diverse candidates.

The company also recognizes that diversity is critical to the future success of Centerra and will strive to enhance its focus on the advancement of all underrepresented groups throughout the company and to foster a culture of equity and inclusion.

Carried Forward — Diversity and Inclusion Commitments

Progress on our diversity goals is monitored by the governance committee.

Diversity and inclusion are among core Aecon values that guide the corporation's thinking, and commitment to such values is championed at the highest levels of the corporation.

Abandoned — One-time Diversity Assessment Initiatives

The company ran its first-ever diversity and inclusion survey.

We recognize that our industry and our company have not yet achieved optimal levels of diversity and inclusion.

$T_{Indigenous_Engagement}$: Indigenous Community Engagement

Introduced — Reconciliation and Consultation Activities

These ambitious goals will allow us to measure our progress in achieving our strategic goals of accelerating the energy transition, advancing Indigenous reconciliation, and supporting the communities we serve, while also supporting our long-term growth.

The company will proceed only when it is safe to do so and Indigenous communities agree to proceed, and when the current restrictions are lifted.

Carried Forward — Indigenous Partnership Arrangements

We have awarded [X] million worth of contracts to Indigenous communities and contractors engaged on the project.

We will continue to partner with Indigenous communities to establish maintenance and operational contracts over our year [X] contract as operator with the Alberta Electric System Operator.

Abandoned — Consent and Project-Access Disputes

These Indigenous communities own and operate businesses that provide services to our project.

It was denied by the Mines and Geoscience Bureau based upon a certification of non-consent issued by the National Commission of Indigenous Peoples, which was in turn based upon the purported failure of Sierra's Philippines associate company, Bunawan Mining Corporation, to secure the free and prior informed consent of the relevant Indigenous cultural communities within the relevant ancestral domains.

$T_{Labour_Relations}$: Labour Relations

Introduced — Collective Bargaining and Labour Disputes

We have no collective bargaining agreements with our employees, and we have not experienced any work stoppages.

None of the company or any of its subsidiaries has engaged in or received notice of any pending or threatened unfair labour practice complaint.

Carried Forward — Employment Standards and Protections

The company will not tolerate any form of forced, compulsory, or child labour in its sphere of influence.

This responsibility takes various forms, including ensuring a decent or living wage for its employees in order to cover the basic needs of a worker and his/her family.

Abandoned — Workforce Management Initiatives

We currently have contracts in place with all of our Canadian unions until [X], providing labour stability and a solid foundation for continued growth.

The company encourages senior management to participate in professional development programs and courses and supports management's commitment to training and developing employees.

Appendix G. S&E Dictionary

Climate Change		
greenhouse gas	ghg emission	ghg intensity
ghg reduction	ghg protocol	ghg inventory
ghg scope	ghg footprint	carbon emission
carbon footprint	carbon capture	carbon cost
co2 emission	co2 equivalent	co2 intensity
co2 reduction	co2e emission	tonne co2e
tonne co2	scope 1	scope 2
scope 3	direct emissions	indirect emission
fugitive emission	methane emission	nitrous oxide emission
nox emission	sox emission	so2 emission
emission reduction	emission intensity	emission reporting
emission factor	emission source	total emission
gross emission	climate change	climate risk
climate policy	climate strategy	climate adaptation
climate resilience	climate mitigation	climate transition
climate impact	transition risk	global warming
greenhouse effect	extreme weather event	energy transition
decarbonization	decarbonisation	kyoto protocol
paris agreement	renewable energy	renewable fuel
clean energy	fossil fuel emission	fossil fuel consumption
energy and carbon	carbon sequestration	ozone layer
ozone deplete substance	perfluorocarbon	sulfur hexafluoride
voc emission		
Ecosystem		
biodiversity	biological diversity	bio diversity
natural ecosystem	aquatic ecosystem	forest ecosystem
terrestrial ecosystem	natural habitat	critical habitat
sensitive habitat	fish habitat	bird habitat
marine habitat	forest habitat	aquatic habitat

riparian habitat	wetland habitat	wildlife habitat
native wildlife	endanger species	threatened species
threatened or endanger	critically endangered	species at risk
native species	invasive species	plant species
native vegetation	natural vegetation	wetland area
wetland protection	wetland loss	riparian area
riparian zone	riparian forest	riparian buffer
adjacent watercourse	watercourse protection	waterway protection
local waterway	boreal forest	rainforest
rain forest	old growth forest	forest conservation
forest management	forest protection	deforestation
reforestation	afforestation	land degradation
soil erosion	soil contamination	soil conservation
soil quality	nature conservation	wildlife conservation
habitat conservation	nature reserve	heritage site
natural heritage	environmental stewardship	marine ecosystem
marine biodiversity	coastal ecosystem	land reclamation
mine reclamation	site reclamation	post mining reclamation
ecological impact	ecological assessment	ecological study
aquatic ecology	forest ecology	fish bearing
fish passage	fish bearing water	migratory species
migratory bird	cultural heritage	archaeological site
heritage resource		

Pollution and Waste		
water management	water quality	water treatment
water contamination	water pollution	water discharge
water conservation	water recycling	water scarcity
water shortage	drinking water	wastewater
wastewater treatment	wastewater discharge	tailings leakage
dam safety	dam breach	dam failure
waste management	waste disposal	waste treatment
hazardous waste	toxic waste	solid waste

mine waste	landfill disposal	landfill site
contaminated soil	contaminated water	contaminated site
site contamination	environmental remediation	site remediation
land remediation	reclamation obligation	reclamation closure
abandonment reclamation	reclamation costs	air pollution
air quality	air emission	dust control
dust management	dust suppression	noise pollution
noise management	noise control	chemical spillage
fuel spillage	oil spill	air pollutant
water pollutant	toxic substance	hazardous substance
acid drainage	acid mine drainage	acid rock drainage
particulate emission	acid generating	non acid generating
potentially acid generating	environmental impact	environmental release
Stakeholder and Community Relations		
indigenous	indigenous people	indigenous community
indigenous rights	indigenous consultation	indigenous land
indigenous territory	first nations	first nation
aboriginal	aboriginal community	aboriginal people
free prior informed consent	free prior and informed consent	duty to consult
treaty right	community	communities
local community	community engagement	community relations
community development	community investment	community benefit
community impact	affected community	host community
local resident	local population	social impact
social license	social licence	social performance
social risk	resettlement	involuntary resettlement
land access	land use	grievance mechanism
grievance procedure	stakeholder engagement	human right
human rights	human right abuse	human right infringement
child labor	child labour	forced labor
forced labour	modern slavery	human trafficking

labour exploitation	social investment	local employment
local hiring	local procurement	employment opportunity
benefit sharing	traditional land	traditional territory
traditional knowledge	rights holder	impact benefit agreement
community agreement	social and environmental	
Occupational Health and Safety		
health and safety	occupational health	occupational health and safety
occupational safety	worker safety	worker health
employee safety	employee health	workplace safety
site safety	mine safety	environmental health and safety
lost time injury	total recordable	safety incident
near miss incident	workplace injury	workplace illness
recordable injury	recordable injuries	worker fatality
mine fatality	worker fatalities	physical injury
reportable injuries	workplace accident	mine accident
personal protective equipment	emergency response	emergency preparedness
emergency evacuation	fire safety	occupational disease
occupational illness	toxic exposure	chemical exposure
occupational exposure	dust exposure	hearing loss
silicosis	safety training	safety management
safety performance	safety culture	safety standard
contractor safety	labour standard	labor standard
labour right	labor right	labour practice
labor practice	freedom of association	collective bargaining
gender equality		
Mining-Specific Environmental Obligations		
environmental permit	environmental assessment	environmental impact assessment
environmental management plan	environmental monitoring	environmental compliance
environmental regulation	environmental legislation	environmental law

environmental approval	environmental license	environmental licence
mine permit	mine closure	mine rehabilitation
closure plan	reclamation plan	reclamation and closure
closure and reclamation	environmental liability	reclamation liability
closure provision	environmental provision	acid mine
environmental fines	environmental damage	geotechnical risk

Appendix H. Pooled BERTopic Analysis

This appendix reports results from a pooled BERTopic model estimated on all three disclosure groups combined (overlapping, AIF-specific, and MD&A-specific sentences). The pooled model identifies 68 topics with a noise share of 40.4%, reflecting the heterogeneity of the combined corpus. The top 20 topics by sentence count are reported below, together with the distribution of overlapping, AIF-specific, and MD&A-specific sentences within each topic.

Pooled BERTopic — Top 20 Topics by Sentence Count

Topic	Label	N	% total	Overlap %	AIF-spec %	MD&A-spec %
0	Environmental Assessment Process	18,765	14.7	22.7	75.2	2.1
1	Water Treatment and Management	5,174	4.1	13.2	85.7	1.1
2	Risk Disclosure	3,973	3.1	41.0	51.8	7.2
3	Occupational Health and Safety	3,084	2.4	16.9	68.8	14.3
4	Environmental Damage and Regulation	2,697	2.1	23.2	74.4	2.4
5	Waste Area Management	2,251	1.8	11.3	88.1	0.6
6	Environmental Costs and Production	2,055	1.6	18.1	79.5	2.4
7	GHG Emissions	1,979	1.5	23.1	75.8	1.1
8	Reclamation and Closure Plan	1,710	1.3	14.6	80.7	4.7
9	Health, Safety and Environment	1,552	1.2	11.3	82.9	5.9
10	Environmental Damage and Liability	1,302	1.0	29.2	67.6	3.2
11	Environmental Regulations	1,228	1.0	25.0	72.2	2.8
12	Closure Plan Approval	1,225	1.0	9.1	86.0	4.9
13	Reclamation Costs	1,171	0.9	24.3	58.0	17.8
14	Community and Local Engagement	979	0.8	20.7	70.9	8.4
15	Community Agreement	910	0.7	16.0	63.4	20.5
16	Land and Price Regulations	866	0.7	22.7	76.1	1.2

Topic	Label	N	% total	Overlap %	AIF-spec %	MD&A-spec %
17	Community Engagement	866	0.7	25.6	65.1	9.2
18	Species and Habitat Protection	854	0.7	11.2	87.7	1.1
19	Social and Community Responsibility	834	0.7	23.6	72.1	4.3

Appendix I. Variable Definitions

Variables	Definitions
RECURRENCE_RATIO	The number of overlapping S&E sentences in an AIF filing divided by the total number of S&E sentences in the filing
LOG_ASSETS	The natural logarithm of total assets at the beginning of the fiscal year.
ROA	Net income divided by beginning-of-year total assets.
LEVERAGE	Total debt (short- and long-term) divided by beginning-of-year total assets.
TURNOVER	The ratio of total revenue over the fiscal year to beginning-of-year total assets.
PPE	Net property, plant, and equipment divided by beginning-of-year total assets.
CAPEX	Capital expenditures divided by beginning-of-year total assets.
MTB	The ratio of market value to book value of total assets, as of the beginning of the fiscal year. Market value of total assets is calculated as the sum of market value of equity, book value of total liabilities, and liquidation value of preferred stock minus deferred tax liabilities.
LOG_PAGES	The natural logarithm of the total number of pages in the AIF filing.
LOG_MDA_SE_COUNT	The natural logarithm of the total number of S&E sentences identified in the MD&A filing.

Appendix J. OLS Regression Results

	RECURRENCE_RATIO
LOG_ASSETS	−0.030***
	(0.005)
ROA	0.015
	(0.013)
LEVERAGE	0.022
	(0.024)
TURNOVER	−0.016
	(0.020)
PPE	−0.015*
	(0.008)
CAPEX	−0.042*
	(0.023)
MTB	0.0004
	(0.002)
LOG_PAGES	−0.136***
	(0.019)
LOG_MDA_ES	0.169***
	(0.007)
Constant	0.884***
	(0.170)
Year FE	Yes
Clustered SE (firm)	Yes
N	2,373
R^2	0.557

Statistical significance is indicated as follows: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.