



[Specialised Masters & MSc

MSc in Big Data & Business Analytics

Master Thesis
Class of 2022

Master's Thesis presented by:

Jeremy Jerome CHIA

Under the supervision of:

Assoc. Prof. Man Ching Gladie LUI

Title:

A Natural Language Processing Analysis of Sustainability Reports and its Relationship with Corporate Financial Performance: Empirical Evidence from Singapore's Listed Companies

Your professional thesis is confidential: ☐ yes ☒ no



[Specialised Masters & MSc

MSc in Big Data & Business Analytics

ESCP benefits from the best international accreditations.
Our European campuses enjoy national recognition.



5 European Higher Ed
accreditations

BERLIN

Heubnerweg 8-10
14059 Berlin, Deutschland

LONDON

527 Finchley Road
London NW3 7BG,
United Kingdom

MADRID

Arroyofresno 1
28035 Madrid, España

PARIS RÉPUBLIQUE

79, av. de la République
75543 Paris Cedex 11, France

PARIS MONTPARNASSE

3, rue Armand Moisant
75015 Paris, France

TURIN

Corso Unione Sovietica, 218 bis
10134 Torino, Italia

WARSAW

c/o Kozminski University
International Relation Office
57/59 Jagiellońska St.
03-301 Warsaw, Poland



European Business Schools
Ranking 2021

Affiliated to



09/22-2020-MS-These

SWORN STATEMENT

I, the undersigned, Jeremy Jerome CHIA,

Student of the Master of Science in Big Data and Business Analytics, certify that I have plagiarised nothing in the paper attached, which means that I am the sole author of each sentence contained in this text. Any other sentence by another author than myself has been enclosed in quotation marks, with explicit indication of its source. I am aware that by breaking this rule, I am violating the fundamental academic principles and subjecting myself to sanctions that will be determined by the disciplinary board.

I also certify that this paper has never been presented in previous studies at ESCP. If this research has been conducted as part of parallel studies, I have to specify it.

The views expressed in this professional thesis are mine alone.



Place: Berlin, Germany

Date: 31 October 2022

A Natural Language Processing Analysis of
Sustainability Reports and its Relationship with
Corporate Financial Performance: Empirical
Evidence from Singapore's Listed Companies

October 2022

Jeremy Jerome CHIA

MSc Big Data and Business Analytics

ESCP Europe Business School

Supervised by: Assoc. Prof. Man Ching Gladie LUI

Abstract

The Singapore Exchange's (SGX) sustainability disclosures have steadily increased since 2016. Sustainability reports were mandated, and climate-related disclosures will soon be required for companies in high-impact industries. This study aims to understand the relationship between Corporate Social Responsibility (CSR) disclosures and financial aspects.

Natural Language Processing (NLP) techniques are employed to process sustainability reports directly to derive the component topics and assess the readability and sentiments of words used. This will be regressed against a profitability measure (the Return on Assets (ROA)), the market value (MV), and financial viability (Zmijewski score). The sample consists of sustainability reports from 30 companies which are component indices of the Straits Times Index from 2015 to 2021.

The results show that specific topics relate to different financial performance indicators. For example, community-related disclosures had a positive relationship with the ROA and MV. Whilst methodology-related disclosures had a positive relationship with MV, resource-related disclosures had a negative relationship with MV. Lastly, governance and climate-related disclosures had a positive relationship with the Zmijewski score, indicating that these factors contributed to a firm's financial viability. Despite this, there was no conclusive evidence that readability nor the sentiments of words used in sustainability reports impacted financial performance.

This study provides a direct deployment of NLP techniques to assess sustainability reports, as opposed to more manual methods used previously. This methodology is an initial assessment in a Singapore context, where companies were found to have the highest and most consistent sustainability disclosure rates in Southeast Asia. The findings also hope to provide a basis for future research to understand CSR factors – how it is perceived by investors and how it impacts financial performances.

Keywords: sustainability, sustainability reports, CSR, disclosures, financial performance, natural language processing, Singapore, listed companies

Acknowledgements

I'm grateful for the support of everyone in helping me bring this piece to fruition. Please allow me to indulge in a few paragraphs of gratitude, nostalgia, and reflection:

My family and loved ones – thank you for supporting my decision to partake in this journey. It's undeniably challenging to grapple with the fact that we must be 10,000 kilometres apart. Yet, I am grateful that our relationship remains strong and close. Thank you for giving me the peace of mind and assurance I need to wholeheartedly embrace life in Paris and Berlin.

Professor Gladie Lui – thank you for being the most encouraging and resourceful supervisor a thesis candidate could have asked for. Our exchanges were heart-warming, and I will cherish them. This thesis is better, thanks to your patience and feedback.

Professors Wei Zhou, Louis-David Benyayer, and the MBD faculty – thank you for unreservedly imparting your knowledge to us. Thank you also, for being there to bounce ideas off with and nurturing our development as data professionals.

Friends back in Singapore and Hong Kong – thank you for keeping me grounded. Thank you also for continuing to involve me in your lives and for allowing me to involve you in mine. Psychologically, having constants helps weather the figurative storms. Special thanks to Augustine for your feedback.

New friends from the Masters, Paris, and Berlin – thank you for being my home away from home. For the lived memories we shared, I am very grateful. Learning about your stories: why and how you are where you are today, was inspiring. In particular – the comforts you traded off to be here. Your experiences have given me much food for thought.

There has been much growth from the last year of being abroad. This thesis is a way of taking stock of this growth professionally, and in hopes of contributing back to the home I grew up in – Singapore, while continuing to explore two fields I am passionate about - accounting and data analytics.

Contents

Abstract	i
Acknowledgements	ii
Table of Figures	vii
Introduction	1
Literature Review	4
Corporate Social Responsibility (CSR) versus Environmental, Social and Governance (ESG)	4
Corporate Social Responsibility	5
Environmental, Social, and Governance	7
The Friedman Doctrine	10
A Rebuke: Strategic Corporate Social Responsibility	11
History of Corporate Social Responsibility	12
Sustainability and the Sustainable Development Goals	16
Sustainability Disclosures	18
Sustainability Reporting Standards	21
Global Reporting Initiative	22
ISSB Exposure Drafts	22
Critique of ISSB's Exposure Drafts	26
Sustainability and Climate Reporting Requirements	28
Sustainability Reports	31
Readability of Sustainability Reports	31
Assurance on Sustainability Reports	32
Theories in Sustainability Disclosure and Financial Performance	33
Legitimacy Theory	33
Institutional Theory	34

Stakeholder Theory	35
Signalling Theory.....	36
Agency Theory.....	37
Resource Theories	37
Transaction Cost Economics.....	38
Conceptual Frameworks.....	38
Factors Considered in Prior Research	40
Dependent Variables	40
Independent Variables.....	41
Natural Language Processing (NLP) in Accounting Research	42
Prior Approaches.....	49
Methodology	51
Research Questions	51
Data Sources.....	52
Description of Samples	52
Dependent Variables	57
Tokenisation and Creating Term-Document Matrices.....	57
Stop Word Removal, Cleaning, and Lemmatization	60
Latent Dirichlet Allocation (LDA).....	62
Selecting the Optimal Number of Topics.....	63
Determining Topics.....	65
Text Readability	66
Text Sentiment	67
Independent Variables.....	72
Models and Approach	74
Analysis on Return on Assets (ROA)	75

Analysis on Market Value (MV).....	75
Analysis on Zmijewski Score (ZJS).....	76
Discussion and Results.....	77
Data Exploration	77
Number of Pages and Words of Sustainability Reports.....	77
Key Topics in Sustainability Reports.....	82
Alignment with Disclosure Themes Required in GRI	83
Distribution of Topics in Sustainability Reports.....	84
Readability of Sustainability Reports.....	87
Sentiments of Sustainability Reports	88
Summary Statistics of Financial Information	92
Interpreting Regression Results	94
Analysis on Return on Assets	94
Analysis on Market Value.....	96
Analysis on Zmijewski Score.....	99
Comparison with Prior Studies	101
Conclusion	102
Limitations and Opportunity for Future Studies	103
References	A-2
Appendix I: Code with Narration.....	B-1
Libraries Required.....	B-1
Loading Sustainability Reports as Corpus	B-1
Summarising Number of Pages.....	B-2
Summarising Number of Words	B-2
Sentiments	B-3
Readability	B-5

Text Cleaning (LDA)	B-6
Topic Modelling (LDA).....	B-7
Regression Analysis	B-9
Appendix II: Additional Stop Words Removed.....	B-10
Appendix III: Evaluation Metrics of “CaoJuan2009” and “Deveaud2014”	B-11
Appendix IV: Readability Scores for Sustainability Reports	B-13

Table of Figures

Table 1: Dimensions of the SDGs proposed by Kang and Kim (2022).....	17
Table 2: Sustainability Disclosure Score for Singapore-listed Companies	20
Table 3: Sustainability Disclosure Requirements by Country	30
Table 4: Sustainability disclosures source by company on SGX's STI.....	56
Table 5: Samples of words removed from the document for analysis	61
Table 6: Interpretation of Readability Ease from the Flesch Reading Ease Score (Flesch, 1948).....	66
Table 7: Lexicons used in this Analysis.....	68
Table 8: Sample Output of Word Sentiments by Lexicon	70
Table 9: Variables used in the analysis.....	74
Table 10: Summary Statistics of Number of Pages in the Sustainability Report by Financial Year	78
Table 11: Summary Statistics of Number of Pages in the Sustainability Report – Overall.....	78
Table 12: Companies with the Minimum and Maximum Number of Pages in their Sustainability Reports by Financial Year.....	79
Table 13: Summary Statistics of Number of Pages in the Sustainability Report by Financial Year	80
Table 14: Summary Statistics of Number of Words in the Sustainability Report – Overall.....	80
Table 15: Companies with the Minimum and Maximum Number of Words in their Sustainability Reports by Financial Year.....	81
Table 16: Proposed Interpretation of Topics.....	83
Table 17: Summary Statistics of Readability Scores for Sustainability Reports	87
Table 18: Summary Statistics of Readability Scores for Sustainability Reports by Financial Year	87
Table 19: Top and Bottom 3 Reports in terms of readability	88
Table 20: Summary Statistics of Sentiment Scores for Sustainability Reports	90
Table 21: Summary Statistics of Sentiment Scores for Sustainability Reports by Financial Year	91

Table 22: Top and Bottom 3 Reports in terms of sentiment scores	91
Table 23: Summary Statistics of ROA for Companies	92
Table 24: Summary Statistics of ROA for Companies by Financial Year	92
Table 25: Summary Statistics of Market Value for Companies	92
Table 26: Summary Statistics of Market Value for Companies by Financial Year...	93
Table 27: Summary Statistics of Zmijewski Score for Companies	93
Table 28: Summary Statistics of Zmijewski Score for Companies by Financial Year	93
Table 29: Results from Multivariate Regression against Return on Assets	95
Table 30: Results from Multivariate Regression against Market Value	98
Table 31: Results from Multivariate Regression against Zmijewski Score	100
Figure 1: Interaction between CSR, ESG, and SDGs (ewenpx3orguk, n.d.)	5
Figure 2: An ESG Journey of a Company (Pérez et al., 2022)	9
Figure 3: Three levels of ambition in ESG for companies (Pérez et al., 2022)	9
Figure 4: SDG nexus representing co-occurrences of predicted SDGs within the Inventory of Business Indicators – the node sizes are proportional to the influence within the network (Matsui et al., 2022)	17
Figure 5: Climate-related Reporting Assessment Framework (GRI and NUS, 2022)	20
Figure 6: Adoption of Various Sustainability Reporting Frameworks for Listed Companies in Reports issued for Financial Periods ending in 2021 (GRI and NUS, 2022)	28
Figure 7: Conceptual Model Linking Corporate Environmental Management and Performance with Firm Value (Reddy and Gordon, 2010)	39
Figure 8: A model of the capital market impacts of corporate social responsibility (Richardson et al., 1999)	39
Figure 9: Rates of positive-negative comments (Kang and Kim, 2022)	45
Figure 10: Thematic distribution of topics within sustainability reports - comparison of six firms from 2011 to 2020 (Kang and Kim, 2022).	46
Figure 11: Mean probability of occurrence of environmental, social, and economic topics in the years 1999 until 2014 (Székely and Vom, 2017).	47

Figure 12: Illustration of Tokenisation	58
Figure 13: Illustration of Creation of Document Term Matrix	58
Figure 14: Top ten (10) words per topic, prior to removing stop words.....	62
Figure 15: Differences between Bag-of-Words and Latent Dirichlet Allocation in text feature extraction.....	63
Figure 16: CaoJuan2009 (Cao et al., 2009) and Deveaud2014 (Deveaud et al., 2014) metric scores by the parameter – number of topics.....	64
Figure 17: Number of Pages of Sustainability Reports by Financial Year	77
Figure 18: Number of Words in Sustainability Reports by Financial Year	79
Figure 19: Number of words versus number of pages in sustainability reports, with a trend line.....	81
Figure 20: Top ten words by identified topic	83
Figure 21: Distribution of Topics across the Years	85
Figure 22: Distribution of topics by year for selected companies	86
Figure 23: Readability Scores by Number of Words in Sustainability Reports	88
Figure 24: Sentiment Score, by Page, in Keppel’s 2015 Sustainability Report.....	89
Figure 25: Average Sentiment Score, with Error Bars representing maximum and minimal values, by Page in Keppel’s 2015 Sustainability Report.....	89
Figure 26: Average Sentiment Score, adjusted for number of pages, by Year	90

Introduction

Non-financial information is increasingly important to stakeholders assessing a company's performance. This morphs the classical view that shareholders exclusively seek to maximise returns. For example, the triple-bottom-line – Elkington (1999) proposed that companies were to focus on people, the planet, and profits. Tsang (1998) observed that several listed companies in Singapore's banking, food and beverages, and hotel industries were making voluntary disclosures on their social corporate responsibility efforts. Such disclosures were often a section in the annual report. These are centred on human resources and community involvement efforts, at 5.3% of all sentences in the annual report. However, environment-related disclosures were scarce – they made up only 0.24% of total sentences. These results represented the broader social focus in Singapore in the 1990s on fostering social cohesion. Since that initial study, other similar studies have revealed statistically significant positive relationships between Corporate Social Responsibility (CSR) disclosures and financial performance (Khavesh et al., 2012).

Sustainability reports have become commonplace for companies to communicate their non-financial performance and initiatives to their stakeholders. These reports are now part of mandatory disclosures for listed companies ("issuers") in Singapore. However, from financial periods ending on or after 31 December 2017, companies listed on the Singapore Exchange (SGX) were required to issue sustainability reports. Practice Note 7.6 stipulated that companies shall describe their sustainability practices on a "comply or explain" basis (Chia and Wee, 2021). The Exchange proposed that firms may use widely adopted frameworks, such as the Global Reporting Initiative (GRI) and the Sustainability Accounting Standards Board (SASB), as a basis for sustainability reporting (Loh et al., 2017). The use of standard frameworks allows for comparability. This also avoids the need for companies to use other costly sustainability initiatives to reduce information asymmetry (Connelly, Ketchen & Slater, 2011).

Following the “Sustainability Reporting Review 2021” by the Task Force on Climate-related Financial Disclosures (TCFD), SGX had strengthened its sustainability reporting requirements. From 1 January 2022, issuers were obligated to perform ‘climate reporting’. This became mandatory for issuers in the (a) financial industry, (b) agriculture, food and forest products industry, (c) energy industry, (d) materials and buildings industry, and (e) transportation industry. A “mandatory” requirement departs from the prior position of “comply or explain”. This shift in regulation signals the Exchange’s initiative to combat climate change and its value on non-financial information such as sustainability reports, allowing stakeholders to assess better the issuer’s financial prospects and quality of management (Singapore Exchange, 2022).

The Singapore Exchange Regulation launched the ESGenome disclosure portal in 2022 (Huang, 2022) as a platform for companies to upload their climate disclosures and for stakeholders to view such information. The SGX RegCo aims to address gaps in sustainability reporting and make it easier for companies to make disclosures. The information required follows the TCFD’s recommendations, and SGX has prescribed 27 additional core metrics (Singapore Exchange, 2021). This initiative will help standardise reported information and potentially improves comparability across reporting firms.

There has been a heightened emphasis on climate change and increasing prominence of social issues in recent years. Regulatory and social developments may result in varied impacts on sustainability disclosures on corporate financial performance (CFP) compared to findings in prior studies. For example, Loh et al., 2017 studied the relationship between CSR disclosures and CFP when SGX instituted a “comply or explain” regime. This was a change from its previous position – encouraged but voluntary disclosures. At this juncture where SGX is strengthening its requirements by making CSR disclosures mandatory for TCFD-identified industries, this study will examine if the relationship between CSR disclosures and CFP for Singapore-listed companies is materially different as compared to other geographies, but also across the time period where sustainability disclosure requirements were strengthened.

There were studies conducted using empirical evidence from listed companies in Singapore (Tsang, 1998; Loh et al., 2017; Wahyuningrum et al., 2021), as well as other jurisdictions – Malaysia (Kasbun et al., 2017), India (Aggarwal, 2013-a; Motwani and Pandya, 2016), Brazil (Ching et al., 2017), Australia (Jones et al., 2007). These studies regressed financial metrics against sustainability indicators at a point in time. In determining inputs for sustainability indicators, these studies used simple, self-developed scoring methods to assess the quantity and quality of disclosures. In the subsequent section, these approaches are discussed.

This study will adapt prior approaches and propose new methods for assessing the relationship between CSR disclosures and CFP. Natural Language Processing (NLP) techniques are adopted to analyse annual reports – topical modelling (Rehurek and Sojka, 2011), readability and sentiment analysis (Pang et al., 2002). These approaches provide a robust yet different dimension of text mining sustainability reports. Input variables from this analysis are regressed against indicators of CFP, such as Return on Equity (ROE) and Return on Assets (ROA), amongst others. The approach will be discussed further in the Methodology section.

This paper will be structured as follows: in the following section, a literature review of theoretical and conceptual frameworks relating to sustainability disclosures and financial performance will be discussed. In addition, prior methodologies used to assess this relationship will be featured. The following section will describe the questions this study attempts to address, the approach adopted for this study, and the data used for the analysis. The results will, after that, be compared to findings in prior studies. Finally, the concluding section summarises the paper and proposes suggestions for stakeholders and future directions for research.

Literature Review

“Capitalism as we know it is dead. A new model of business is taking its place, driven by values, ethics, and taking care of employees”.

Marc Benioff, Salesforce Chief (Edgecliffe-Johnson and Mooney, 2019)

Corporate Social Responsibility (CSR) versus Environmental, Social and Governance (ESG)

Whilst CSR and ESG are two terms which appear to be used interchangeably; there are differences in what these encompass. The difference is important to understand as the discussion on sustainability reports continues, as sustainability reports in recent years are presented in line with frameworks that prescribe, or guide, disclosure of the latter – ESG efforts.

The emergence of these two terms hints at the difference between CSR and ESG in business strategy. CSR stemmed from a market failure in the 19th and 20th centuries, whereas ESG is an integration of various fragments of initiatives from 2005. CSR is driven by a reactionary force, whereas ESG is a proactive force driven by a desire to do better. Later, the United Nations (UN) agreed on the Sustainable Development Goals (SDG), discussed in a later section. Ewenpx3orguk (n.d.) proposes a Venn diagram – Figure 1 – which attempts to explain the intersection between these concepts.

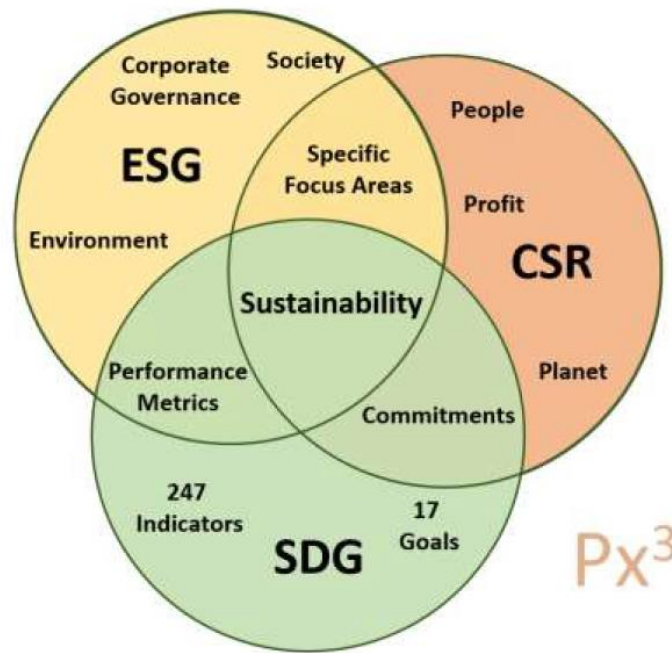


Figure 1: Interaction between CSR, ESG, and SDGs (ewenpx3orguk, n.d.)

This subsection attempts to provide references to academic and professional definitions of the first two terms – CSR and ESG.

Corporate Social Responsibility

Corporate social responsibility emerged during the industrial revolution but became more established as a concept through corporate philanthropy. Despite numerous efforts to achieve a clear and unbiased definition of Corporate Social Responsibility, there is still some confusion about how it should be defined (Dahlsrud, 2008). Several definitions of Corporate Social Responsibility (CSR) include:

- A duty for companies to pursue and make appropriate decisions or actions (Bowen, 1953).
- The **obligations** of businessmen to pursue those policies, to make those decisions, or to follow those lines of action which are desirable in terms of the objectives and values of our society" (Bowen 1953; Ealls, 1956; Selekman, 1959).

- Businessmen's decisions and actions are taken for reasons at least partially beyond the firm's direct economic or technical interest (Davis, 1960).
- The social responsibility of business encompasses the economic, legal, ethical and discretionary expectations that society has of organisations at a given point in time (Carroll, 1979).
- A concept whereby companies **integrate** social and environmental concerns in their business operations and interact with their stakeholders voluntarily (European Commission, 2001).
- Corporations implement voluntary action to pursue their mission, fulfilling their perceived obligations to stakeholders, including employees, communities, the environment, and society (Coombs and Holladay, 2012).
- The moment in which corporations reflect their concerns about social and global issues on their activities, even if some activities may not strictly be related to the core of their business (Trapp, 2012).
- A form of business regulation that guides companies to operate socially responsibly. Core subjects listed in CSR include human rights, fair labour practices, the environment, fair operating practices, consumer issues, and community involvement and development (Green Business Bureau, 2021).

Between these definitions, one commonality to highlight was the fulfilment of perceived obligations. The assumption is that CSR adversely affects corporate performance regardless of behavioural motivations. This implies a need for continual evaluation of agendas against an evolving set of expectations considering societal structures and practices in the present (L'Etang et al., 2011). CSR is a critical decision of corporates that involves stakeholders and sustainability (Dahlsrud, 2008). The emphasis is on the premise that there are activities that meet social expectations while consuming certain costs.

Some studies have argued that reasons for companies to take on more social responsibility are motivated by a desire to have more profitability than competitors with less social responsibility (Orlitzky et al., 2003; Vogel, 2005).

The conflicting view on CSR harbours three themes: (i) violating the obligation to shareholders - which, explained later through the Friedman Doctrine, is to maximise profits, (ii) covering wrongdoing, and (iii) creating false dichotomies (Freeman & Dmytriiev, 2017). CSR has also been critiqued for its 'voluntary transparency regime' – which can be counterintuitive. This is because CSR may be a decoy or a legitimising function, as corporations can easily claim to be transparent, thus 'inoculating the public against negative accounts of corporate behaviour' (Nadesan, 2011).

Environmental, Social, and Governance

ESG is viewed as a way of categorising CSR initiatives. The term is self-explanatory through its components – Environmental, Social, and Governance matters. From one perspective, the term can be understood as a measurement concept, a subset of CSR initiatives.

In 2005, the term ESG first appeared at the "Who Cares Win" meeting involving investors, analysts, and government agencies (Kim and Park, 2021). It was a paradigm shift. Albeit, even before that, there were similar concepts that pointed toward the same thing - demands for ESG management are gradually increasing as the scope of stakeholders surrounding companies is more comprehensive than in the past (Berman et al., 1999). The 2005 meeting had the effect of formalising these initiatives and this premise into a term.

Jin (2022) observed that ESG encourages investors and companies for long-term interest and returns. It is driven by external factors – rather than market failure; it is investors' interest and concerns about whether a company is sustainable. For example, ESG disclosures exhibit how the company responds to risks that may undermine its long-term viability.

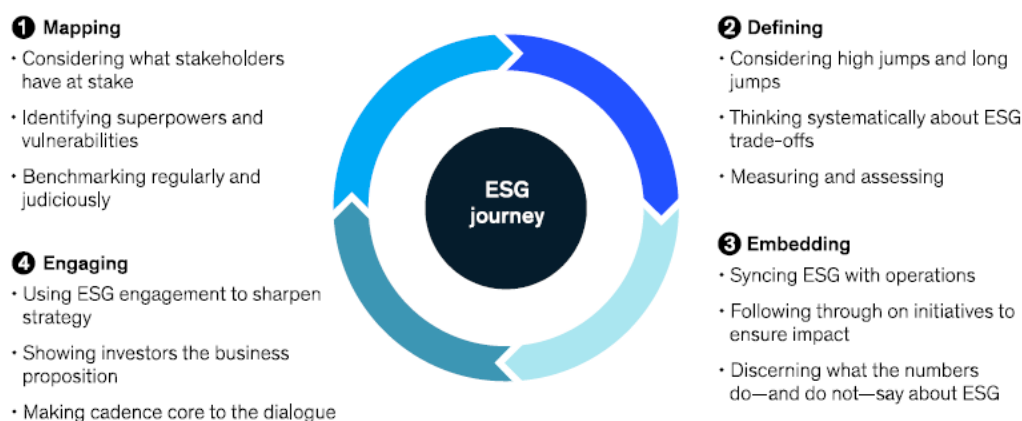
Some attempts to discuss ESG include the following:

- The concept of CSR - although confusing - could be framed as a combination of three main components: Environmental, Social and Governance (Hansmann et al., 2012). These components are exactly what the ESG acronym stands for, thus identifying a possible measurement tool for a concept that still seems abstract.
- Different sectors have different priorities, and ESG lacks any single definition of what should be included but linking ESG to compliance reporting provides some opportunity for science-based metrics (Rodgers et al., 2013).
- ESG is a corporate responsibility encompassing not only traditional financial factors but non-financial factors, such as environment, society, and governance (Kim, 2021). ESG management, on the other hand, emphasises the importance of non-financial factors such as environment, society, and governance in evaluating corporate valuation.
- KRX (2021) described ESG as a concept that encompasses factors related to the environment, society, and governance structure that can affect a company's ability to implement strategies and increase its value.
- MSCI (2021) observed that ESG investment is a term often used synonymously with sustainable investment, social responsibility investment, mission-related investment, or screening.
- ESG sets specific criteria to define sustainable environmental, social, and governance systems. ESG points to a specific set of criteria that removes sustainability's ambiguity (Green Business Bureau, 2021).

Not all aspects of 'E', 'S', and 'G' are priorities for all companies. It is unrealistic to expect that companies do not have to make hard trade-offs within and among ESG dimensions or that they can lead on every topic. This hints at why in the initial phases of sustainability disclosure, companies chose only to disclose a limited range of topics – on issues on which they had information or were of immediate interest to their stakeholders in the area in which they operate (Pérez et al., 2022). Figure 2 and Figure

3 drives the idea that ESG is a process, not an outcome, because of the dynamicity of stakeholders' expectations.

ESG is a process, not an outcome.



McKinsey
& Company

Figure 2: An ESG Journey of a Company (Pérez et al., 2022)

	Common practice	'Next level' practice
Minimum practice <ul style="list-style-type: none"> Risk-mitigation and “do no harm” measures • React to trends affecting industry and business • Address external vulnerabilities • Donate resources (financial, in-kind, volunteer) • Meet and report baseline standards • Pledge to minimal commitment levels 	<ul style="list-style-type: none"> Substantive efforts, mostly outside the core business • Track major trends affecting the business with contingency plans in place • Use strengths to deliver increased value across specific ESG goals and metrics • Comply with voluntary industry standards and perform above industry average • Create a comprehensive sustainability policy • Implement inclusive HR practices • Run strategic, high-impact philanthropic programs • Engage with stakeholder groups to understand what matters 	<ul style="list-style-type: none"> Full integration of ESG into strategy and operations • Leverage “superpowers” to move sector standards • Increase social impact via innovation, market, and customer choices • View ESG as a differentiator and core to overall strategy • Link clearly articulated leadership areas (“win the game”) with purpose • Embed ESG in capital and resource allocation • Tie ESG to employee incentives and evaluations • Improve sustainability outcomes internally and externally • Ensure that ESG disclosures cover company’s full operations

Figure 3: Three levels of ambition in ESG for companies (Pérez et al., 2022)

Research into CSR has demonstrated the relationship between corporate environmental, social, and economic performance (Brogi and Lagasio, 2019; Chowdhury et al., 2019; Taliento et al., 2019). Empirical evidence suggests that companies focusing on social, economic, and environmental issues will gain a competitive advantage and have a credible reputation and public opinion (Pechancová et al., 2019; Fatoki, 2019). These three factors are interconnected and tackling these factors in tandem assures long-term sustainability (Pirnea et al., 2011).

Pérez et al. (2022) suggest that a company's decision-making and shift towards sustainability can be approached with either 'high jumps' or 'long jumps'. The former refers to levels companies must reach to meet its arbitrary ESG bar – the gap between where it is and where society expects it to be. The latter refers to one or two ESG areas where the company can take a leadership role and affect other players in its ecosystem and beyond. Pérez et al. (2022) posit that these two elements are necessary for companies looking to build a competitive advantage.

The Friedman Doctrine

The traditional view of a company's obligation to its shareholders is described in the Friedman Doctrine. Friedman said that the essential purpose of a company is to maximise profits (Friedman, 1970).

He believed that CSR and ESG activities should be minimised as much as possible. Weidenbaum and Vogt (1987) outlined that a company's decision-making for the environment, society and governance, not for shareholders' interests, is a factor that hinders corporate performance by incurring opportunity costs. From the viewpoint of shareholder capitalism – if a manager, who is an agent of shareholders, consumes corporate resources for CSR instead of shareholders' profit, it undermines shareholder interests (Friedman, 1970). He believed that managers take advantage of corporate social responsibility as an agenda for their social and political careers. This is a stark contrast to the current view outlined in the next section that discusses the developments in societies and companies' perspectives of CSR.

Separately, he held the view that the State, and not companies, has a moral and social responsibility toward society.

A Rebuke: Strategic Corporate Social Responsibility

Chandler (2016) held that firms should focus on what they can do best to optimise and create value rather than maximise profits. This starts with incorporating the CSR perspective into a company's strategic planning process and corporate culture. A difference in the premise is that the company's actions are directly related to its core operations and that companies need to be responsive to their stakeholders' needs. The firm should aim to optimise the value created rather than profits. The company, therefore, will pass from a short-term perspective to mid- and long-term planning (Chandler and Werther, 2013).

The development of the business environment's view on CSR is a journey and is discussed in the next section.

History of Corporate Social Responsibility

Global developments have shaped the philosophy and premise behind CSR. This section attempts to outline historical developments that have led to the understanding that institutions and stakeholders face CSR in its present form in the 2020s. Global events, and the efficacy of CSR initiatives, have shifted the perception of CSR from being viewed as a trade-off to one that provides a strategic advantage for companies. An overarching view was that CSR remained an essential element of corporate response to varied social needs (Jamali, 2008) and was taken as a method of taking on social responsibility in consideration of the impact of its operations on stakeholders and a way of gaining society's trust (Panait et al., 2014). The critical driver of CSR development is events that shaped the world and changing expectations of stakeholders.

The board of directors also has a significant piece in the morphing of CSR. The board is responsible for corporate strategies, culture, and governance and approves major decisions (Basel Committee, 2015). In turn, their decisions significantly influence the success or failure of the company (García-Sánchez et al., 2018). Empirical evidence suggests that various characteristics – gender diversity, independence, and the presence of an ESG committee – influence CSR performance (Kyaw et al., 2017). The shapers of CSR are boards that are more diverse, independent, and naturally more focused on ESG with an ESG management committee. Stakeholder demands are shifting, which can dramatically affect competitive dynamics (Pérez, 2022).

Because financing is a crucial component which drives business development – the financial industry was expected to play a leading role in allowing other industries to conduct ESG management through the way it structures and requests information for its financial services and investment products.

These developments are highlighted by period:

- **The 1940s**: During World War II, companies' growth and political instability led corporations to be seen as institutions with social relevance, and the first discussions on corporate social responsibility were coming to life (Heald, 1970).
- **The 1960s**: Particularly in the United States, this decade marked years of protests, sit-ins, and walk-outs by students and rallies – most notably, the anti-apartheid and civil rights movements, which demonstrated how people's power could dramatically affect businesses. Big companies played a role similar to decision-making organisations as their actions began to influence the society they operate within.

CSR was a response to labour-management conflict. This was exacerbated by a change in the artisan's work model towards mass production in tandem with the industrial revolution. The capitalist model neglected behaviour that caused corporate human and labour rights infringement (Jenkins, 2009). In addition, there have been calls for responsible actions by companies concerning social aspects (Gomez-Carrasco et al., 2016).

- **The 1970s**: marked an era of "managing corporate social responsibility" (Carroll, 2015). The first Earth Day was celebrated in 1970 in response to the 1969 oil spill in Santa Barbara. The scale and dramatic event changed society's faith in companies and regulatory frameworks. This later influenced corporations' behaviour and responsibilities towards society (Clarke and Hemphill, 2002; Spezio, 2018).

In academia, a framework was also proposed to rate firms as 'outstanding', 'honourable mention', or 'worst' by exploring the relationship between CSR and financial performance (Moskowitz, 1972; Moskowitz, 1975).

- **The 1980s**: Concerns about environmental influence from human behaviour began to emerge (Salvioli, 2000). The goal of CSR policy morphed to focus on improving companies' image and reputation and securing social justification to practice (Carrol, 2008).
- **The 1990s**: The last decade of the 20th century marked the beginning of the institutionalisation of CSR. This trend was sustained by globalisation and the increased prominence of multinational firms. An awareness of supply chains and international labour practices became the focus, and CSR focused on a combination of profit, social good, and anti-slavery commitments. Albeit, there was much resistance as these trends began to take flight.

The emerging view was that corporate ethical behaviour would increase corporate competitiveness by establishing sustainable and productive relationships with stakeholders (Jones, 1995). However, there was a consensus that a company that performs well with eco-friendly policies is a resource in creating a competitive advantage (Russo and Fouts, 1997).

- **The 2000s**: The turn of the century also came with a turn of events for companies and CSR. The acceleration of globalisation and the complexity of business relations with various social and interest groups brought about more reforms to the company model (Cornelius et al., 2007; Shnayder et al., 2016). The 2008 Financial Crisis caused corporate policies to become the subject of immense public scrutiny (Kaplan, 2020), fast-tracking the process to institutionalise ESG practices. Corporate policies were changing in response to the immense public interest (Smith, 2011). The fulfilment obligations of firms to their stakeholders were taking shape – to have the intention of minimising harm and maximising the long-run beneficial impact of the firm on society.

CSR started to become part of business strategy, with a view that doing so boosts the competitiveness and efficiency of resource utilisation. "Doing good" became a competitive advantage as potential investors began to view ESG

initiatives as essential factors (Vibert, 2019). In other words, the shift in social responsibility transformed "CSR from being a minimal commitment to becoming a strategic necessity, which can translate into a sustainable competitive advantage." (Warther & Chandler, 2005). This is the decade that saw CSR integration into business strategies (Acro-Castro et al., 2020).

- **The 2010s**: A shift in the view of a corporation – Porter and Kramar (2011) propose that – "the purpose of the corporation must be redefined as creating shared value". There is an acceleration of the institutionalisation of CSR as well. Several standard-setting bodies emerged and proposed how disclosures should be performed. At an institutional level, proposals for model norms for corporate governance in many countries were encouraged. Sustainability became a joint CSR commitment – and was frequently linked to SDGs.

External shocks such as the COVID-19 pandemic and the war in Ukraine have shown that companies find it hard to move unless they have an ESG framework derived from and deliberately advance their strategy.

Sustainability and the Sustainable Development Goals

The UN 2015 set up 17 Sustainable Development Goals (SDGs) intended to be achieved by 2030¹. This superseded the Millennium Development Goals, which ended in 2015 — branded as "The future we want" – the UNSDGs was adopted at the UN Conference on Sustainable Development in Rio de Janeiro in June 2012.

There are several attempts to help stakeholders relate to and understand the themes within the SDGs. The uninitiated stakeholder with limited knowledge of SDGs typically encounters difficulties mapping their local challenges and activities onto the broader SDG context. SDGs' interlinkages are complex (Bowen et al., 2017). The Open-Source SDG developed a holistic SDG ontology by coupling a conventional framework and an SDG multi-label classification system by linking regression and topic models (Pukelis et al., 2020). Kang and Kim (2022) also proposed that the 17 SDGs can be condensed into six (6) categories representing different human needs dimensions. These are summarised in Table 1.

Category	Sub-Category	Goal
Social	Equity	SDG 4: Quality Education
		SDG 5: Gender Equality
		SDG 10: Reduced Inequality
	Social Development	SDG 11: Sustainable Cities and Communities
		SDG 16: Peace, Justice and Strong Institutions
		SDG 17: Partnership for the Goals
Economic	Life	SDG 1: No Poverty
		SDG 2: Zero Hunger
		SDG 3: Good Health and Well-Being
	Economic and Technological Development	SDG 8: Decent Work and Economic Growth
		SDG 9: Industry, Innovation and Infrastructure

¹ Taken through Resolution No. [A/RES/70/1](#) of 2015.

Environmental	Resources	SDG 6: Clean Water and Sanitation
		SDG 7: Affordable and Clean Energy
		SDG 12: Responsible Consumption and Production
		SDG 14: Life Below Water
	Environments	SDG 13: Climate Action
		SDG 15: Life on Land

Table 1: Dimensions of the SDGs proposed by Kang and Kim (2022)

Matsui et al. (2022) have also proposed another way of understanding the connection between SDGs by visualising the SDGs nexus by analysing the co-occurrences of predicted SDGs multi-labels with the Inventory of Business Indicators from the SDG compass. There are strong associations between SDG 1 and SDG 10, which are social goals, and between SDG 7, 11, 13, 14 and 15 – which are environmental goals pinned around sustainable cities. This demonstrates that regardless of which areas of the SDGs companies focus on, the goals are not isolated and have interconnected effects on other impact areas.

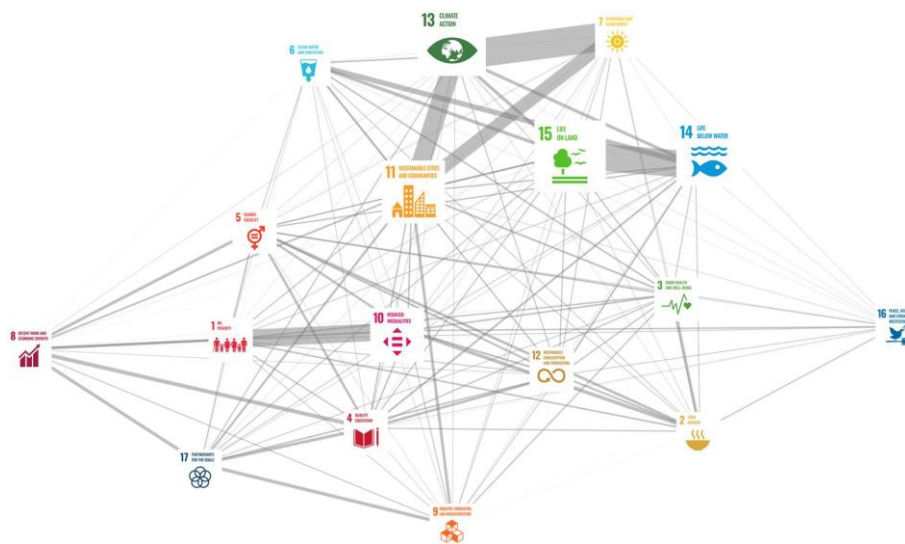


Figure 4: SDG nexus representing co-occurrences of predicted SDGs within the Inventory of Business Indicators – the node sizes are proportional to the influence within the network (Matsui et al., 2022)

Sustainability Disclosures

There are several mediums for corporations to perform disclosures. Notwithstanding the medium, such communication is typically structured with an intent to persuade (Nwagbara and Belal, 2019). Therefore, a communication strategy could take the path of least resistance to reach the audience the message is meant for. The traditional format is through periodic publications – in the form of the annual report and, more recently, the sustainability report for institutional investors. The other format is through social media or internet channels, which allow for higher frequencies and more formats (Parasharya, 2022). This study focuses on the former – formal reporting channels through reports as this has a more historical basis for comparison.

Initial disclosures were more unstructured compared to the current form. As a rebuke to the Friedman Doctrine, Danone led a change by adopting an internal accounting policy that reflected carbon emission costs to Earnings per Share (EPS) to create awareness about the greenhouse gases emitted from food manufacturing. This was one way sustainability disclosures took shape – with companies disclosing topics that might have been easier or advantageous for them to disclose, rather than an entire comprehensive suite of disclosures as we now see today. Consistency was essential to realising comparability.

Parasharya (2022) observed, in 2007, that conglomerates like ExxonMobil and PepsiCo did not have distinct sections for sustainability reports – such information was nested within broader themes (e.g., “Purpose”). This is similar to Singapore-listed companies. One of the first adopters of a separate Sustainability Report is City Development Limited, which presented its first Sustainability Report in 2012. Companies also started to disclose sustainability initiatives as a section of their annual report. Other listed companies gradually started having separate reports, which became mandatory in 2017.

The number of organisations that publish information on their sustainability practices has grown steadily (Kolk, 2004). Corporations have been incorporating ESG metrics

as a standard practice into their sustainability reporting and more significant CSR initiatives (Kaplan, 2020; Tett, 2021). However, some researchers have focused on the frequency of reporting and other high-level information to gain insights into the general development of sustainability reporting (Kolk, 2003; Kolk, 2004). This was prior to Natural Language Processing (NLP) techniques gaining traction and feasibility in this research field.

Disclosures can be done using ESG Metrics. These include Greenhouse Gas (GHG) emission metrics, waste, employee health and safety, compliance, and governance-related metrics (Green Business Bureau, 2021).

The TCFD's 2019 status report observed that only 29% of companies made relevant climate disclosures. However, there was insufficient decision-useful climate-related financial information (FSB TCFD, 2019). This is despite prior efforts to release a set of recommendations to help structure and formalise companies' sustainability and climate risk reporting (TCFD, 2017). In subsequent reviews, the TCFD was increasingly prescriptive on the expected disclosures. Their 2021 review had prompted SGX strengthening its sustainability reporting requirements to require 'climate reporting'.

When ESG is core to the business model, reporting on ESG becomes part of the ordinary course of doing business. In so, external shocks are less likely to present an undue burden on ESG reporting (Pérez, 2022).

GRI and NUS (2022) made attempts to analyse how companies made disclosures based on seven key areas: (i) reporting framework, (ii) materiality, (iii) risks and opportunities, (iv) governance, (v) strategy, (vi) targets, and (vii) performance. Figure 5 outlines the climate-related reporting assessment framework.



Figure 5: Climate-related Reporting Assessment Framework (GRI and NUS, 2022)

The study noted that Singapore-listed companies excelled in consistently tracking and disclosing across historical periods, making retrospective comparison possible. **This makes Singapore-listed companies a favourable choice for the analysis.** Notably, Singapore-listed companies performed reasonably well in the disclosure scoring study by GRI and NUS (2022). The results are reproduced in Table 2 below.

Company Name	Ticker	In STI?	Score
City Developments Limited	C09	Yes	96.4
Singapore Telecommunications Limited	Z74	No	96.4
United Overseas Bank Limited	U11	Yes	92.9
Wilmar International Limited	F34	Yes	89.3
Keppel Corporation Limited	BN4	Yes	82.1
Ascott Residence Trust	HMN	No	78.6
Fraser and Neave Limited	F99	No	78.6
Sembcorp Marine Limited	S51	No	78.6
SATS Limited	S58	Yes	75
Singapore Airlines Limited	C6L	Yes	75

Table 2: Sustainability Disclosure Score for Singapore-listed Companies

Several studies posit that the disclosure of social, economic, and environmental issues in the sustainability report has several effects. First explored in a discussion on signalling and stakeholder theory, disclosures were acknowledged as a way to increase

transparency (Kaymak and Bektas, 2017). As the legitimacy theory explains, disclosures increase accountability, reputation (Aguilera-Caracuel et al., 2018), and awareness of environmental and social practices (Chang et al., 2017).

A challenge is to convince stakeholder groups that the narratives presented in such disclosures are more than merely ‘greenwashing’ (Bartlett, 2011). The hidden motives of self-interest can be attributed to ‘psychological egoism’, fuelled by an assumption that altruism solely for the benefit of others is extraordinary (Parasharya, 2022).

Sustainability Reporting Standards

Following financial shocks in the first two decades of the 21st century, companies have sought ways to “redeem themselves” - ESG standards are a crucial development in improving reputation. ESG reporting is more standardised – there is consensus amongst corporations regarding the form of practice (Waddock and Googins, 2011). The development of standards is a step towards formalising the consensus.

Sustainability disclosures currently consist of a mix of mandatory and voluntary initiatives. This contrasts with financial reporting, which is guided by the International Financial Reporting Standards (IFRS) with a principle-based approach and guidance on the basis for recognition, measurement, and prescriptions for disclosures for comparability. SGX, similar to other stock exchanges, has proposed that firms may use widely adopted frameworks, such as the GRI and the Sustainability Accounting Standards Board (SASB), as a basis for sustainability reporting (Loh et al., 2017).

The standards are meant to guide businesses to give a snapshot of their business impact across critical areas – environment, social, and corporate governance. Standardised reporting methods, such as GRI, summarise quantitative and qualitative information for easy disclosure and improved transparency in screening investments. ESG reporting helps investors avoid companies that may pose a greater financial risk due to environmental performance or other social or government practices (Green Business Bureau, 2021).

Global Reporting Initiative

The GRI is currently the most widely adopted sustainability reporting standard (GRI and NUS, 2022), as evidenced in Figure 6. It was formed in 1997 with the support of the United Nations Environment Programme. The GRI's standards have a broad stakeholder focus – investors, employees, suppliers, customers, and the community.

The standards are presented in three main buckets – (i) Universal Standards, (ii) Sector Standards, and (iii) Topic Standards. Whereas there are three Universal Standards which are applicable to all companies that set minimal expectations for disclosures about the reporting organisation and disclosures and guidance about the organisation's material topics.

ISSB Exposure Drafts

The International Sustainability Standards Board (ISSB), backed by the IFRS Foundation, released two exposure drafts for sustainability reporting standards. This is meant to harmonising existing standards – e.g., the GRI. The preface outlined the board's impetus and premise for the project:

“Sustainability-related risk and opportunities arise from an entity's dependencies on resources and its impacts on resources, and the relationships the entity maintains that may be positively or negatively affected by those impacts and dependencies.” (ISSB, 2022a)

IFRS S1 guides sustainability-related financial information, whereas IFRS S2 provides climate-related disclosures. The impetus for doing so was outlined separately:

“The degree and type of exposure of the effects of climate-related risks and opportunities on the assessment of enterprise value are likely to vary by sector, geography, and entity. Users want to understand the specific response.” (ISSB, 2022b)

The primary users of financial reporting need more consistent, complete, comparable, and verifiable sustainability-related financial information to help them assess an entity's enterprise value. Whereas the definition of enterprise value is set out in IFRS S1 paragraph 5:

Enterprise value reflects expectations of the amount, timing and certainty of future cash flow over the short, medium and long term and the value of those cash flow in the light of the entity's risk profile and its access to finance and cost of capital. Information that is essential for assessing an entity's enterprise value includes information provided by the entity in its financial statements and sustainability-related financial information.

Standardisation promotes an extent of comparability. However, there is variance due to the difference in standards that can be chosen. The standards are based on the principle of materiality – which is subjective to the firm's interpretation. This results in disclosures that are often very heterogeneous and lacks a structure in the subjects they cover (Luccioni et al., 2020).

GRI and NUS (2022) proposed a climate-reporting framework with seven main disclosure areas in response to climate-disclosure mandates. This framework purports to cover key disclosure areas required for stakeholders to understand the reports in a structured manner. There are significant overlaps between this proposal and the requirements prescribed in the IFRS Foundation's exposure draft for Sustainability standards (IFRS S1, IFRS S2). The areas identified by GRI and NUS (2022) for disclosure are:

1. Reporting framework (e.g. GRI, SASB, SDG, TCFD),
2. Materiality,
3. Risk and opportunities,
4. Governance (particularly how the board committee is involved),

5. Strategy (how the organisation incorporates climate-related disclosures into its strategy),
6. Targets (provision of metrics and time frames),
7. Performance (particularly in terms of Scope 1, 2, and 3 emissions)

In the ISSB's exposure drafts, the four broad categories of disclosures are:

- **Governance (IFRS S1:11a, 12-13)**: Providing disclosures on processes, controls, and procedures used to monitor and manage sustainability-related risks and opportunities.
- **Strategy (IFRS S1:11b, 14-24)**: Company's strategy for addressing significant sustainability-related risks and opportunities and the effects of significant sustainability-related risks and opportunities on the financial position, financial performance, and cash flows.
- **Risk management (IFRS S1:11c, 25-26)**: The process by which sustainability-related risks and opportunities are identified, assessed, and managed.
- **Metrics and targets (IFRS S1:11d, 27-35)**: How the entity measures, monitors, and manages its significant sustainability-related risks and opportunities.

In addition, there are prescriptions on the features of information required to be disclosed that extend slightly beyond the categories identified above:

- **Related information**: a connection between various sustainability-related risks and opportunities and how it relates to information in the general-purpose financial statements.
 - IFRS S1:44 explains the combined effects of sustainability-related risks and opportunities and the entity's strategy and related metrics and targets on its financial position, financial performance, and cash flows over the short, medium and long term. For example, an entity may face decreasing product demand because of consumer preferences for low-carbon alternatives. It may need to explain how its strategic response, such as closing a major factory, affects its workforce and communities, as well as the effect of the closure on the useful lives of assets and impairment assessments.
- **Materiality**: Issues for misstating or obscuring that information could reasonably be expected to influence decisions that the primary users of general-purpose financial reporting make based on that reporting.
- **Comparative information**: to disclose in respect of the previous period.
- **Location of Information**: it should be part of its general-purpose financial reporting, integrated with the current FS disclosures and not separate.

The text in IFRS S2 is mainly similar to IFRS S1, except that it is worded for climate change. There is a greater specification on the details – for example, IFRS S2:21(a) under metrics make specific mention of Greenhouse Gases (GHG) Scope 1, 2, 3 emissions. IFRS S2 also stipulates that information should be disclosed separately for groups – in terms of making a distinction between subsidiaries and reporting on the activities of its associates and joint ventures. This puts an added interest in the parent

owners to take an interest in its holdings where it has significant influence – creating more avenues of influence to steer corporate strategy towards ESG.

Critique of ISSB's Exposure Drafts

Public accounting bodies have responded with critiques on the exposure drafts. The ISSB laid out questions to solicit feedback as part of the exercise. The general feedback from these bodies was that the ambiguity in ISSB's proposed standards should be reduced and that more terms need to be defined tightly. It was highlighted that standardisation between different standard-setting bodies should be a goal to be pursued.

Strickland (2022) highlighted that creating a consistent global sustainability taxonomy across jurisdictions is essential. The American Institute of Certified Public Accountants (AICPA) expressed concern that the ISSB and European Financial Reporting Advisory Group (EFRAG) might create competing standards. The EFRAG's standards were observed to be more prescriptive than principle-based – similar to the divergence between the General Accepted Accounting Principles in the United States (perspective) and the International Financial Reporting Standards (principle-based). The TCFD structure was adopted in ISSB's standards but not in the EFRAG's proposals. Harding (2022) noted that where differences in key definitions – (i) verifiability and understandability, (ii) materiality, and (iii) accounting for enterprise value creation. Chartered Institute of Management Accountants (CIMA) observed that while the EU's objective was to "build on what exists and seek as much alignment as possible" was virtuous (Harding, 2022), there is some way to put this into practice.

On materiality, CIMA's critique was that there were differences in the definition of materiality in the ISSB's exposure draft vis-à-vis EFRAG's guidance, which was an unnecessary variation (Harding, 2022).

A separate critique said that more specific information on the materiality of sustainability information is needed to guard against obscuring the disclosures' usefulness (Strickland, 2022). However, the critiques found it necessary to have more guidance on 'significant' should be interpreted to inform better what should be reported. For example, Strickland (2022) pointed to the Committee of Sponsoring Organisations of the Treadway Commission (COSO) framework to report risk.

Ultimately, these proposals aim to make sustainability information reported through the standards valuable and comparable. This is particularly useful when comparing the sustainability performance across different industries and geographies – in dimensions most likely to cause incompatibility. The challenge remains in harmonising and finding the right balance between prescription and a principle-based approach.

Sustainability and Climate Reporting Requirements

Since 2016, there has been a push from stock exchange regulatory bodies in Southeast Asia to mandate sustainability disclosures. Regulators across SEA have different timelines and varying requirements for sustainability disclosures. At the point of writing, there is no common sustainability reporting framework across the Association of South-East Asian Nations (ASEAN) member states (Loh et al., 2018). However, a commonality between regulators was a consensus on the GRI as a prescribed reporting framework. Figure 6 shows the percentage adoption of different sustainability reporting frameworks for listed companies in the respective countries' stock exchanges for reports issued for financial periods ending in 2021.

	GRI	IIRC	SASB	SDG	TCFD
Indonesia	93%	4%	16%	93%	5%
Malaysia	73%	35%	11%	74%	19%
Philippines	82%	17%	35%	86%	38%
Singapore	99%	8%	10%	65%	18%
Thailand	89%	13%	10%	95%	27%
Vietnam	65%	8%	2%	42%	0%

Figure 6: Adoption of Various Sustainability Reporting Frameworks for Listed Companies in Reports issued for Financial Periods ending in 2021 (GRI and NUS, 2022)

Compared to other ASEAN countries, Singapore excelled in tracking and disclosing metrics across historical periods. This aids trend analysis and makes Singapore-listed companies a suitable candidate for the analysis. In reports issued for financial periods ending in 2021, 99% of Singapore-listed companies reported with the GRI. This highlights the effectiveness of regulatory power in influencing disclosures. Singapore has also substantially improved climate-related reporting from 2019 to 2021 (Loh and Tang, 2021).

Regulators had also sought to portray themselves as a supporter of the TCFD² recommendations – given its prominence as a grouping established by the G20 Financial Stability Board. An expectation is that with a regulatory push and a more robust mandate, more companies would incorporate disclosures recommended by TCFD. For example, Fang et al. (2022) compared regulatory efforts across six ASEAN member states – Indonesia, Malaysia, the Philippines, Singapore, Thailand, and Vietnam. This is summarised in Table 3:

Country	Disclosure requirements for publicly-listed companies	Reporting components required	Support of TCFD	Other collaborations
Indonesia	From 2020, to publish sustainability reporting through POJK51 / POJK 03/2017.	Sustainability governance and performance, including: <ul style="list-style-type: none"> • Economic • Green environment • Social aspects 	June 2021, IDX announced its ambition to support sustainability in Indonesia's capital markets.	In 2021, collaboration with GRI – will provide reporting frameworks for the country's UN SDG commitments.
Malaysia	Required as a listing rule. Advisory guidance through a Sustainability Reporting Guide in 2015 and 2018.	Material economic, environmental, and social risks and opportunities. No mandate on the choice of sustainability framework.	Encourage disclosure using TCFD recommendations.	
Philippines	From 2019, the “comply or explain” approach. From 2023, mandated to comply with sustainability reporting guidelines. Plans to extend this to non-listed companies.	Built on globally accepted standards and frameworks – GRI, IIRC Integrated Reporting Framework, SASB Sustainability Accounting Standards. (SEC TH Website)		

² The Task Force on Climate Related Financial Disclosures (TCFD) was established in December 2015 by the G20 Financial Stability Board, and holds considerable influence in companies' disclosures.

Singapore	<p>From 2017, required to disclosure five primary components. (Walker, 2021)</p> <p>From 2022, more detailed disclosures. (Aravindan, 2021)</p> <p>From 2022, subject the sustainability reporting process to internal or external review.</p> <p>From 2022, mandatory board directors training.</p>	<p>From 2017, requires five primary components of:</p> <ol style="list-style-type: none"> 1. Reporting framework, 2. Materiality assessment, 3. Policy, practices, and performance/reporting, 4. Target setting, 5. Statement by the board <p>From 2022, additional requirements include:</p> <ul style="list-style-type: none"> • Board diversity disclosures • Proposed core ESG factors (27 factors) 	<p>Mandatory and ‘comply or explain’ disclosures based on TCFD recommendations.</p> <p>To require mandatory climate reporting based on TCFD recommendations.</p>	
Thailand	<p>From 2021, mandatory to report via Form 56-1 One Report within three months from the publication of the financial report.</p>	<p>Disclose information on at least the following matters:</p> <p>Overall policy Operation and reporting Business impacts and responsibility Social and environmental activities</p> <p>No prescription on frameworks is required.</p>	<p>In Nov 2021, it announced support for TCFD.</p>	
Vietnam	<p>From 2016, requires disclosures (Circular No. 155/2015/TT-BTC,) of the social and environmental consequences of their activities, and their social commitments in their annual report.</p>	<ol style="list-style-type: none"> 1. Environmental impact (e.g. GHG emissions) 2. Raw materials management 3. Energy and water consumption 4. Compliance with environmental protection laws 5. Employees policies 6. Report on responsibility for the local community <p>Report on green capital market activities.</p> <p>Encouraged to apply globally accepted reporting and disclosure standards.</p>		

Table 3: Sustainability Disclosure Requirements by Country

Sustainability Reports

These sustainability disclosures have culminated in sustainability reports. Especially for listed companies, communication strategies are aimed at engaging with the constant public scrutiny per the basic standards of ESG reporting (Parasharya, 2022).

Readability of Sustainability Reports

Townsend et al. (2010) observed that the primary motivation behind readers' use of sustainability reports is to inform (i) decisions on the use of an organisation's products or services and (ii) investment or divestment decisions. Instead, the GRI proposes a tenant of 'balance' – that reports should ideally be composed of a balance between positive and negative news (GRI, 2013).

The sustainability report is a specialised document that is widely read, given the critical information, but its readers do not read it extensively. This can be attributed to its difficult readability. For example, the average reader reads three (3) reports, while the 95th percentile of readers read only between 10 and 20 reports per year (Smeuninx et al., 2020).

While sustainability reports have much in common with the notably difficult-to-read genre of corporate financial reporting, financial information typically targets a specialised readership of investors and analysts who are likely better equipped to deal with its complexity (Courtis, 1995; Courtis, 1998; Stanton and Stanton, 2002; Li, 2008a). The average person will struggle to decode sustainability reports (Farewell et al., 2014). Sustainability reporting is a developing phenomenon, and its current consumers may not be as well equipped to read them. The readability of these reports is potentially an accessibility barrier for the information in these reports to inform the stakeholders it seeks to reach. Companies should, therefore, work harder to choose simple language (Farewell et al., 2014). Similarly, regulatory bodies should work towards the same.

Interestingly, it was noted in Abu Bakar (2011) that the communications' readability for Malaysian firms deteriorates as the company performance does, supported by the obfuscation hypothesis of Courtis (1998) and Rutherford (2003).

Parasharya (2022) observed that one of the key and recurring insights is the immensity of the nominalisations, generalisations, and abstractions, which can be observed in NIKE, PepsiCo, and ExxonMobil's websites. The frequent use of 'we', 'our', 'commitment' and other market-oriented buzzwords on innovation and transformation makes for a genericised reading that leaves a lot unsaid, an assessment like the one made by Ledin and Machin (2018).

Assurance on Sustainability Reports

Assurance is an essential component in building trust. However, more attention must be paid to auditing in the spirit of attaining comparable and reliable sustainability information free from material misstatements, similar to financial information.

The International Auditing and Assurance Standards Board had intended to draft an auditing standard in response to the ISSB's exposure drafts of IFRS S1 and S2. In its comment letter, it commented that the exposure drafts lacked clarity about the meaning, scope, or application of certain key concepts, which may, in addition to other challenges, create a completeness challenge for assurance practitioners (Brewer, 2022).

In the context of reporting standards, the 'ability for assurance' means whether they are designed in the manner that results in sustainability information on which an assurance engagement can be performed (Brewer, 2022). However, the challenge remains that in IFRS S1 and IFRS S2, multiple criterion levels may be used in the recognition, measurement, presentation, and disclosure of sustainability information. This makes providing reasonable assurance a significant challenge.

The prevailing standards for assurance on non-financial information include ISA 720 on "The Auditor's Responsibilities Relating to Other Information" and ISA 805 on

“Special Considerations – Audits of Single Financial Statements and Specific Elements, Accounts or Items of a Financial Statement”. A new standard is needed to address sustainability-related information as the current ISAs relate to audits of financial information predominately, with little guidance on non-financial information, which sustainability- and climate-related information falls under.

Theories in Sustainability Disclosure and Financial Performance

Sustainability reporting is a complex phenomenon that a single theory cannot explain (Cormier et al., 2005; Tagesson et al., 2009). Therefore, unifying all empirical findings within one theoretical framework remains a challenge. Barnett (2007) remarked that we now “understand the effects of isolated pieces of the overall puzzle, *ceteris paribus*, but the dots remain unconnected through any theoretical framework”. Therefore, it is necessary to maintain an openness that there may be mediator or moderator variables (e.g. a firm’s investment in intangible assets) which define the relationships between sustainability disclosure and financial performance (Blanco et al., 2012) – because there are inevitably still discrepancies among studies (Litzky et al., 2003).

Nevertheless, a combination of several theories – legitimacy, institutional, stakeholder, signalling, and agency theory – aids an understanding of the theoretical relationship between CSR disclosures and financial performance.

Legitimacy Theory

The legitimacy theory suggests that it is necessary to achieve society’s approval for the company to survive. Companies need to act in congruence with society to uphold their business activities. (O’Donovan, 2002). Companies are encouraged to seek legitimacy through disclosures (Cho and Patten, 2007), some even being pressured by their stakeholders to make significant corporate social responsibility efforts (Brammar et al., 2007).

Firms interested in being legitimised by society will take an interest in reporting environmental information. CSR disclosures influence stakeholders' and, eventually, society's perceptions about the company, resulting in higher firm value and better financial performance (Hooghiemstra, 2002).

Companies adopt several procedures and policies, such as voluntarily disclosing environmental information, in order to enhance their perceived legitimacy (DiMaggio and Powell, 1883; Scott, 1995; Aguilera and Jackson, 2003; Delmas and Toffel, 2004). These motivations are exhibited through the topics companies choose to report and the extent to which they report.

Trust is a key element of building legitimacy. Capriotti (2011) and Hovering (2018) observed that communicative transparency and ongoing dialogue with stakeholders about the corporation's 'responsible' behaviour is necessary. Integrated reporting, in the early stages of sustainability disclosures, was the 'common language of transparency' – CSR was viewed as having a role in 'shaping the socio-political struggle over the meaning of sustainable development and the power-knowledge relationships that are taking place in its name (Livesey and Kearins, 2002).

Institutional Theory

The institutional theory posits that structures that include norms and routines are established as authoritative guidelines for social behaviour. These structures become socially acceptable and legitimate within the environment's organisational field. Companies within similar institutional operating environments are coerced into behaving similarly. For example, Pucheta-Martinez et al. (2018) found that firms operating in Anglo-American corporate governance systems tend to report environmental information voluntarily. Ownership structure and type affect intellectual capital disclosure (Li et al., 2008b), and this observation can be similarly extended to environmental reporting policies. Jensen and Berg (2012) and Fasan et al. (2016) have demonstrated that a country's institutional features drive integrated reporting and quality.

This study focuses on firms listed on the Singapore Stock Exchange in the Straits Times Index. These primarily consist of firms domiciled in Singapore and Hong Kong. These firms are likely to operate with an Anglo-American corporate governance system – where there is high-ownership dispersion, a financial system oriented towards markets, strong investor protection from a common law legal system, shareholder orientation and the scarce presence of dominant owners in firms (Pucheta-Martinez et al., 2018). The expectation is that these firms will report environmental information more voluntarily.

Pucheta-Martinez et al. (2018) find several factors are significant in affecting environmental reporting: (i) **ownership dispersion** encourages firms to publish environmental information to address the risk of information asymmetry, (ii) more **financing in the capital market** encourages transparency through public disclosures, and (iii) the type of industry. However, they found that investor protection was not a statistically significant factor. These observations reflect institutional theory's observations and have similar elements to stakeholder and signalling theory.

Stakeholder Theory

A common approach to analysing firms' corporate disclosure is the Stakeholder theory. (Huang and Kung, 2010). This suggests that firms accede to the needs of stakeholders – more than merely the needs of shareholders (Ruf et al., 2001) – because of stakeholders' expectations of a firm. For example, it is insufficient for managers to focus exclusively on the perceived needs of shareholders (McWilliams et al., 2006). They (firms) will be influenced not only by agents with strong power and dependency (shareholders, employees, or investors, among others - considered internal stakeholders), but also by external stakeholders, such as social and environmental activities, professional critics, the media or the press (Mitchell et al., 1997). When applied to a firm's commitment to social activities, stakeholder theory supports a firm's investment in CSR to enhance its relationship with its customers, employees, and shareholders (Rodgers et al., 2013).

In this regard, Pucheta-Martinez et al. (2018) posit that stakeholders expect firms to voluntarily disclose CSR matters such as environmental and social information because it helps them mitigate information asymmetries and capital costs. Several studies have concluded that environmental, social, and governance factors are influenced by planning mechanisms and the management's control over the ability of a firm to create value for stakeholders (Jamali, 2008). The recent view is that doing so would improve relationships between stakeholders and companies in the long run and protect the expectations and interests of stakeholders (Cupertino et al., 2019).

Signalling Theory

Signalling theory underpins the notion that an agent (company executives) conveys information to the principal (stakeholders). In sustainability, investors and consumers have difficulties determining which firms are genuinely committed to sustainability (Connelly et al., 2011). CSR activities and ESG management can also be seen as a type of signal companies send to their stakeholders (Backhaus et al., 2002). Companies that disclose environmental issues are engaged in proactive environmental strategy. They seek to indicate underlying qualities and build a corporate reputation by showcasing the execution of CSR processes (Varda, 2014).

Because external stakeholders are not necessarily privy to the internal operations of a company, they rely on signals sent by companies. The effect of the signal depends on how stakeholders interpret the signals emitted by the company through CSR activities and ESG management. These companies have a higher corporate reputation and attractiveness. People can react to a firm's CSR investment by seeking employment with the firm instead of just purchasing its products. (Greening and Turban, 2000). This, in turn, improves corporate competitiveness (Greening and Turban, 1997). These effects may indirectly lead to better financial performance as well.

There is an incentive to inform principals by voluntarily disclosing more, as positive signals are more appealing (Clarkson et al., 2008). Varda (2014) posited that there are incentives for companies with better results to disseminate those results more

efficiently, and potentially vice versa. There are also other spill-over effects. For example, employees of companies with excellent ESG management performance show less work neglect and absenteeism, which positively influences labour productivity (Flammer and Luo, 2017).

Sentiment analysis of sustainability reports may reveal insights into the relationship between the sentiments expressed in these reports and financial performance.

Agency Theory

In tandem with agency theory, voluntary disclosure of firms, mainly on social and environmental aspects, is a means to reduce current or potential agency costs that may occur in the form of legislation and regulation (Galani et al., 2012). This would close the gap between principals' and their agents' misaligned objectives. By way of reducing information asymmetry, sustainability reporting would have positive effects on a firm's perception.

Resource Theories

Perceived ethics underlies one of the key metrics that determine the success or failure of any organisation, trust. This is a form of organisational capital and is a resource as necessary to a company as cash reserves (ewenpx3orguk, n.d.).

The slack resource theory posited that corporates possessing more surplus could actively pursue CSR activities and ESG management than companies that do not (Waddock and Graves, 1997). However, extra resources generated by a company's performance are a decisive factor in influencing the company to pursue CSR activities and ESG management more actively (Artiach et al. 2010).

The resource-based theory supports the slack resource theory by furthering that a company's internal resources decide the source of corporate competitive advantage (McWilliams and Siegel, 2011). CSR activities help managers develop better skills,

and firms develop intangible assets (e.g. brand name), which contribute to better financial performance (Russo and Fouts, 1997).

On the other hand, meeting stakeholders' demands can be seen as a strategic investment – a resource (Ruf et al., 2001). By investing in such a strategy, organisations develop assets that are valuable, rare, and non-substitutable, such as leadership and a positive social reputation – leading to competitive advantage and potentially higher return (Luo and Bhattacharya, 2006).

The theory explains that CSR activities act as a company's competitive advantage.

Transaction Cost Economics

This idea was mooted by Jones (1995) – that firms would try to satisfy stakeholders' needs to minimise potential transaction costs (Williamson, 1985). It was explained that firms with good CSR perceptions have low-cost implicit claims (from other stakeholders), whereas those with poor CSR perceptions are more likely to face high-cost explicit claims from shareholders and debt holders (Cornell and Shapiro, 1987; Peloza, 2006) for higher risks involved.

Conceptual Frameworks

Conceptual frameworks expanded on the classical theoretical foundations in providing a contextual understanding of this phenomenon. For example, Richardson et al. (1999) propose a model to explain the capital market impacts of CSR, as shown in Figure 8. In addition, Reddy and Gordon (2010) present a conceptual framework to explain the causality of corporate environmental efforts with firm value, as shown in Figure 7.

These conceptual frameworks suggest that CSR efforts will impact firm value. For example, a key intermediary in the process is signalling through disclosures. Communicating their efforts through disclosures would affect perceptions of risk, performance, and management, leading to effects on CFP.

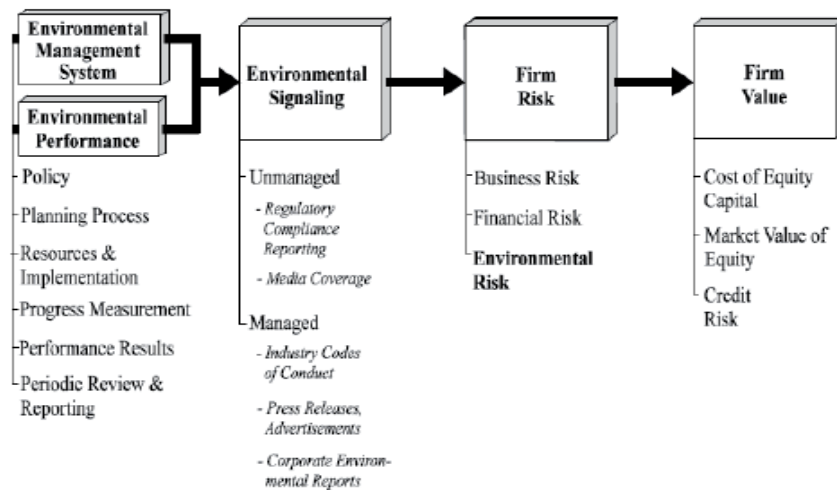


Figure 7: Conceptual Model Linking Corporate Environmental Management and Performance with Firm Value (Reddy and Gordon, 2010)

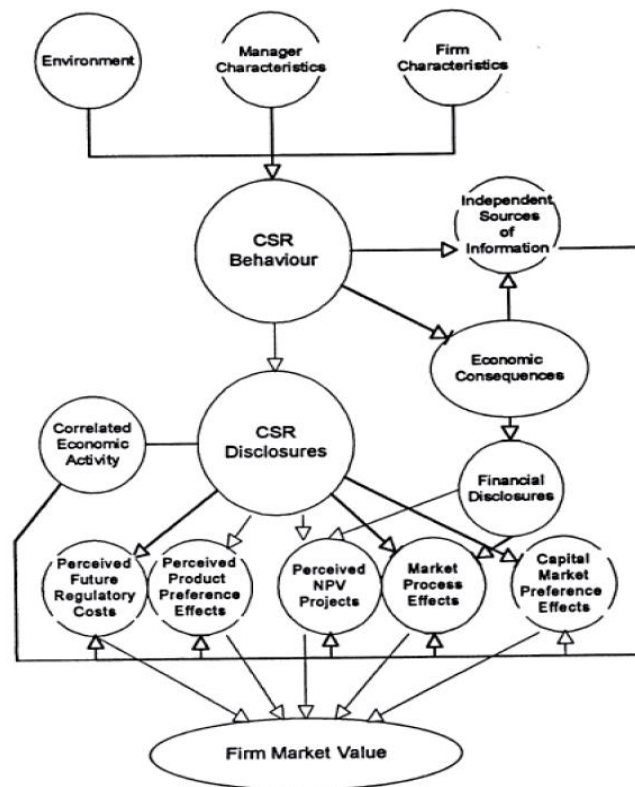


Figure 8: A model of the capital market impacts of corporate social responsibility (Richardson et al., 1999)

Factors Considered in Prior Research

Dependent Variables

The factors used to determine sustainability disclosures and performance were varied. For example, Loh et al. (2017) used a measurement scheme developed by the ASEAN CSR Network, which proposes 23 criteria in the scheme over governance, economic, environmental, and social indicators. Wahyuningrum et al. (2021) used an abridged form of content analysis to assign scores based on the number of disclosures – the number of sentences and the number of pages. Kasbun et al. (2016) similarly used the number of sentences to indicate the quantity of activity reported. Aggarwal (2013-a) used scores extracted from private databases provided by Bloomberg and Thomson Reuters Eikon.

On the other hand, Meng et al. (2013) assigned values on a scale from 0 to 3 to specify the extent to which the information was present in the reports. For example, Dangelico and Pontrandolfo (2010) manually processed reports in the sample and used binary variables (1, 0) to indicate the presence or absence of environmental information in the sustainability reports. Henry (2006) went one step further by employing keyword counts (accomplished with Diction 5.0 software) and thesaurus-based keyword analysis. Cöster et al. (2020) considered a sustainability reporting maturity grid to assess the quality of sustainability reports. Most of these metrics can often be superficial and do not extensively examine the text of such CSR disclosures.

ESG ratings were also one of the options considered. In that, ratings were downloaded from providers such as Bloomberg or Thomson Reuters and used as dependent variables. Agencies actively differentiate themselves by focusing on specific dimensions of the ESG paradigm to establish their own identity in the market. (Boulash et al., 2013). Companies have also responded to these assessments differently, depending on the diversity of their external and internal environments (Delmas and Toffel, 2008; Philippe and Durand, 2011). An issue with these is in comparability and objectivity – since different agencies use different methodologies.

As new NLP techniques emerged, subsequent studies considered more extensive textual analyses. For example, in Pucheta-Martínez et al. (2018), categorical principal component analysis (CATPCA) and partial triadic analysis (PTA) were used to obtain a numerical value for each of the environmental issues of companies that were included in the research. This method attempts to extensively examine the content of the sustainability reports, and similar methods shall be adopted for this study.

Increasingly, studies have used the entire report's text ('corpus') as dependent variables. The mechanism of how this works is explained in the methodology section subsequently. For example, Brown and Tucker (2011) and Qiu et al. (2014) used term frequency-inverse document frequency (tf-idf) term weighting in their analysis.

Some studies supplement their analysis by using outputs of NLP analysis of the corpus as dependent variables. For example, this could include the readability score or sentiment score. Balakrishnan et al. (2010) used the tf-idf weight, the Gunning Fog index for readability and sentiment as explanatory variables for firm performance. Engelberg (2008) used typed-dependency parsing (using the Stanford parser), enhanced with tone analysis.

Independent Variables

Financial metrics used were based on generally-available information. For example, figures were retrieved from audited financial reports of the same year. Standard metrics used to assess performance were the Return on Assets (ROA), Return on Equity (ROE), and Profit before Tax (PBT), employed in several studies (Aggarwal, 2013-a; Ching et al., 2017; Kasbun et al., 2016). Jones et al., 2007 examined the effect on abnormal stock returns, Rodgers et al. (2013) and Rodgers et al. (2017) used the Zmijewski score (Zmijewski, 1984) as a proxy for financial viability. The Zmijewski score is modelled using the Return on Assets, Financial Leverage, and Liquidity, and the coefficients had positive signs, and the relationships were statistically significant.

Although market valuation is not within the scope of discussion of this study, it is worth noting that Tobin's Q is the metric choice in Rodgers et al. (2013) for market valuation. Tobin's Q is the market's valuation of a firm relative to its assets-in-place – the sum of the market value of equity and the book value of debt, scaled by the book value of assets. Finally, it provides a size-adjusted valuation of the company – and can be adopted for this purpose.

Natural Language Processing (NLP) in Accounting Research

NLP can be defined as a theoretically motivated range of computational techniques for analysing and representing naturally occurring texts at one or more levels of linguistic analysis to achieve human-like language processing for a range of tasks or applications (Liddy, 2001). NLP-driven research emerged in the late-1990s (Fisher et al., 2016).

The researchers' original focus was on manual text analysis, but the computerised options allowed for more machine learning (ML)-centric applications (Fisher et al., 2016). Back et al. (2001) implied previously that prospective research could provide clarity on how NLP-based tools could steer stakeholders to consider previously overlooked qualitative (i.e., textual) data in tandem with traditionally analysed quantitative (i.e., financial) data. NLP techniques have allowed for a more efficient analysis of financial reports. Given the substantial volume of information emerging from companies' disclosure, NLP techniques have allowed researchers and analysts to reduce the time and effort required to identify climate-relevant disclosures from such reports. Moreover, the trend toward analysing larger data sets has been enabled by the efficiency and enhanced processing capabilities of NLP applications combined with ML and artificial intelligence (O' Leary, 2013).

Earlier works had only considered a limited number of reports because of the lack of processing power (Székely and Vom, 2017). Some studies used qualitative content analysis techniques to provide an overview of various organisations' reporting practices (Feundlieb et al., 2013). Manual text analysis established that financial statement disclosures, including management discussion and analysis, could predict

future firm performance (Singhvi, 1968; Ingram & Frazier, 1980, Kothari et al., 2009, Lawrence, 2011).

Whereas, increasingly, other studies examined the content of sustainability reports more quantitatively through text-mining techniques, a subset of NLP, focusing on the frequency of keywords related to sustainability practices (Modapothala, 2009; Modapothala, 2010; Liew, 2014). Text mining studies reaffirmed the association between disclosures and future firm performance (Merkley, 2011, 2014; Campbell et al., 2014).

Despite the momentum, developing NLP techniques in finance has proven challenging because of the specialised language used (Luccioni et al., 2020). For example, ‘short’ and ‘bear’ do not have the same meanings in finance as in general society. Most NLP dictionaries are built for generic text. For example, the Natural Language Toolkit designed by Bird et al., 2009 features the Valence Aware Dictionary and sEntiment Reasoner (VADER), a pre-trained dictionary for social media text (Hutto et al., 2014). Nonetheless, the developments thus far have allowed for a more extensive and objective analysis of textual data, which will be considered in this analysis.

In accounting research, many methods have been employed to aid with the analysis of textual information. This include, but is not limited to, the following:

1. The question-answering technique in Lan et al. (2019) identifies the presence of specific information, i.e. the answers, based on a set of prescribed questions. Luccioni et al. (2020) employ this technique to analyse financial reports and identify climate-relevant sections.
2. Machine translation is used in Ott et al. (2018) to compare texts from companies domiciled in different locations. Translating texts to a single language (e.g. French to English, German to English) allows for comparing texts across more jurisdictions.

3. Natural language inference is used in Conneau et al. (2017) to determine whether a “hypothesis” is an entailment, a contradiction, or if it is undetermined given a “premise”. This can aid the association.
4. A Bidirectional Encoder Representations from Transformers (BERT) for domain-specific applications was used in Devlin et al. (2018) for language understanding.
5. Sentiment Analysis was used in Mohammad (2016) and is a common technique for detecting the text's effectual states, emotions, and valence. This was also performed in an exploratory manner by Kang and Kim (2022).

It was found that the ratio of positive and negative comments did not differ significantly by company. Companies are generally consistent in their deployment of positive and negative words over time, keeping their messaging consistent under the signalling theory. For example, a recent study demonstrated that Indonesian companies appear to use a choice of words that contain more positive sentiments (75%) vis-à-vis negative sentiments (25%) (Harymawan et al., 2020).

However, when specific events occur, the positive-negative ratio rate increases. This is supported by the “obfuscation hypothesis”, which posits that companies will make unfavourable news more difficult to decode (Courtis, 1998; Rutherford, 2003). This may also result in more positive language or the deployment of attractive imagery for window-dressing and greenwashing (Hrasky, 2012; Boiral, 2013). This is exhibited in Figure 9 below.

6. A Thematic Analysis, as adopted in Székely and Vom (2017) and Kang and Kim (2022), aims to summarise and identify the themes in the text based on Latent Dirichlet Allocation (LDA), a form of dimensionality reduction, to identify themes and the distribution in the corpus (Blei, 2012). For example, Székely and Vom (2017) used LDA to identify (i) companies’ sustainability

practices and their development over time, (ii) the coverage of economic, environmental, and social aspects in sustainability reports, and (iii) the differences in sustainability reporting and practices among specific sectors.

The technicalities of LDA will be discussed in the methodology section.

An analysis of the thematic structure of sustainability reports by Kang and Kim (2022) showed that the distribution of themes is similar across time. However, across companies of different natures and industries, this varies. Following the six dimensions proposed and documented in Table 1 above, the findings are presented in Figure 10.

Székely and Vom (2017) made observations through thematic analysis and documented the mean probability of occurrence of environmental, social, and economic topics in topics released from 1999 to 2014. It is seen that economic sustainability topics are of increasing importance for organisations, especially after the economic shocks of 2008-2009 (Freundlieb, 2013). Also, from 2010 onwards, environmental, social, and economic sustainability topics were more equally distributed. These trends are presented in Figure 11.

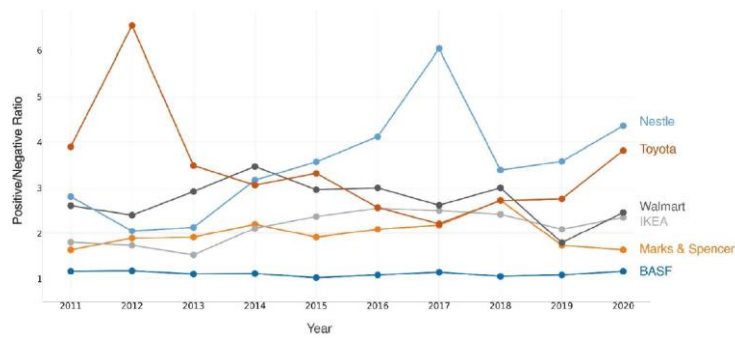


Figure 7. Rates of positive/negative comments.

Figure 9: Rates of positive-negative comments (Kang and Kim, 2022)

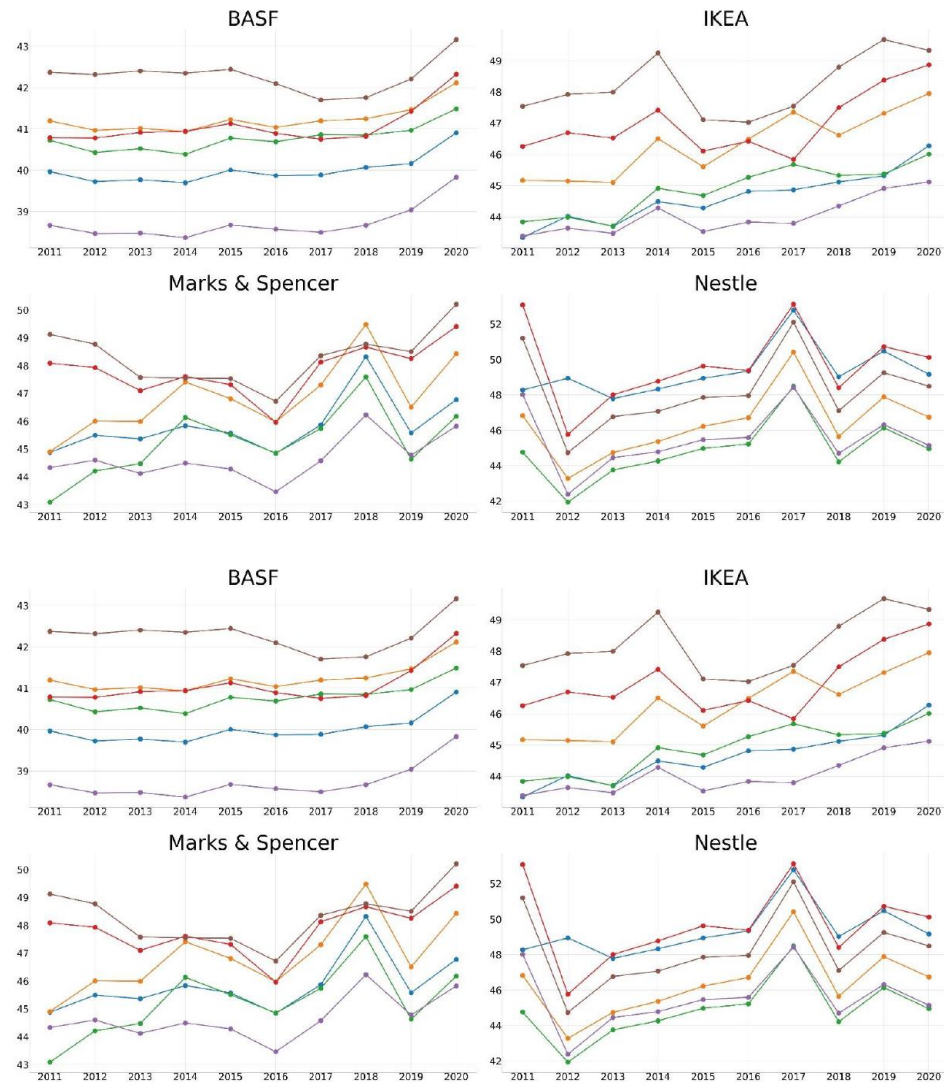


Figure 10: Thematic distribution of topics within sustainability reports - comparison of six firms from 2011 to 2020 (Kang and Kim, 2022).

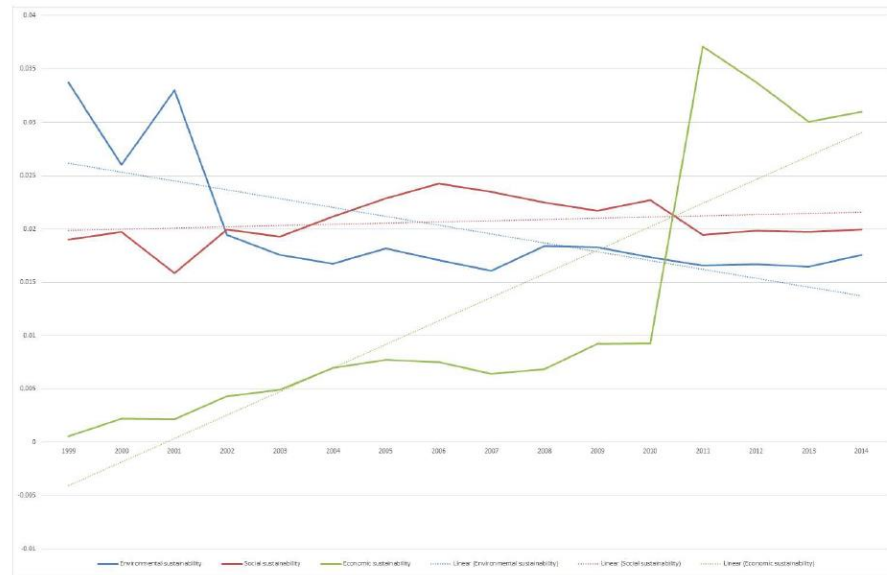


Figure 11: Mean probability of occurrence of environmental, social, and economic topics in the years 1999 until 2014 (Székely and Vom, 2017).

7. A readability assessment was employed in Smeuninx et al. (2020). There are three commonly used readability indices:
 1. The Flesch Reading Ease Score (Flesh, 1948) – comparing the corpus readability with that of other genres – most commonly used,
 2. The Flesh-Kincaid Grade Level score (Kincaid et al., 1975) – quantifies the years of education that the text requires of the reader, and
 3. The Gunning Fox Index (Gunning, 1952; revised in Bogert, 1985) – distils a grade-level measure similar to the Flesh-Kincaid but places a stronger emphasis on the ratio of polysyllabic (“complex”) to mono- or disyllabic words present in the text.

Regarding the metrics above, Butler and Keselj (2009) suggested that prospective analysis could use readability analysis to identify ‘companies that produce more easily read annual reports, indicating that it is a safer investment.

In addition, the readability of corporate reports and disclosures may be a product of management motivation to obfuscate poor firm performance (Butler & Keselj, 2009) and even fraud (Goel et al., 2010).

It was observed that in companies and reporting years with better economic performance, (i) the linguistical parse trees were shallower, indicating less complexity in the text, and (ii) there was the more active language used, which signals more internal attribution (Smeuninx et al., 2020). Internal attribution of results (i.e. the internal locus of control) is invoked when there is a favourable financial performance, while a poor performance invites external attribution of results. On the contrary, passive structures allow for obfuscation, enabling report writings to conceal agency (Rutherford, 2003).

This is similarly confirmed in prior literature. Manual text analysis established that annual report readability is a direct predictor of firm performance, with poor readability associated with poor performance and vice versa (Smith and Taffler, 1992a, 1992b; Clatworthy and Jones, 2001, 2003, 2006). Studies employing basic text mining confirmed this phenomenon (Othman et al., 2012; De Franco et al., 2012; Loughran and McDonald, 2014). NLP, combined with Flesch, Flesch-Kincaid, and Fog index scores to assess readability, enabled the team to forecast upcoming annual stock performance with over 60% accuracy (Butler and Keselj, 2009).

Prior Approaches

Prior studies on this topic used empirical evidence to test the statistical significance of the relationship between CSR disclosures and CFP. These have minute variances. Some studies chose to regress listed firms at a point in time to determine if more disclosures in the year had a positive relationship with financial performance (Kasbun et al., 2017; Aggarwal, 2013-a; Motwani and Pandya, 2016; Ching et al., 2017; Jones et al., 2007). Others conducted a longitudinal study comparing companies' performance across years (Tsang, 1998; Wahyuningrum et al., 2021). In longitudinal studies, the authors sought to examine if an improvement in CSR disclosures resulted in a similar improvement in CFP.

Studies examined used regression analysis in testing the statistical significance. Regression, through the Partial Least Squares (PLS)³ method, was preferred by Rodgers et al. (2013) and Rodgers et al. (2007) for its utility in understanding structural models involving multiple latent constructs with multiple indicators. The PLS has less stringent assumptions with minimal restrictions on distributions and sample size (Chin et al., 2003) and can test complex constructs with both reflective and formative factors (Chin, 1998). Loh et al. (2017) used the Ohlson model (Ohlson, 1995) as a baseline model. They added proposed sustainability indices to determine the improved significance of the models to determine the explanatory effect of sustainability variables included. Several others opted to use each financial indicator as a dependent variable iteratively and sustainability factors as an independent variable. Firm size, a function of the total assets, was also considered in the model. Engelberg (2008) used regression analysis to demonstrate that the tone of the company's announcements is positively correlated ($p < 0.01$) with future firm performance ($p < 0.01$).

³ PLS is a latent structural equation modelling technique that allows the researcher to test the relationship within measures and the hypothesised relationship between measures simultaneously (Lohmoller, 1989).

Some studies took classification-based approaches, using a pre-trained model. Li (2010) used a pre-trained Navie-Bayesian machine learning classification algorithm to show that changes in the tone of management discussion and analysis were positively correlated with future firm performance ($p < 0.01$). Balakrishnan et al. (2010) and Qiu et al. (2014) used an SVM classifier algorithm to investigate the correlation between corporate disclosures and future firm performance ($p < 0.05$, $p < 0.001$, respectively). Brown and Tucker (2011) used a vector space model classifier to identify financial reactions to management discussion and analysis modifications ($p < 0.10$).

To take an exploratory angle, studies considered unsupervised methods. For example, Bao and Datta (2014) employed an unsupervised LDA topic model to learn the topics in the corpus before using a kNN classification algorithm to analyse textual disclosures.

Some studies sought to distinguish results by industry and other characteristics. For example, Ching et al. (2017) looked only at companies that were determined to have the best level of disclosure – for example, companies classified as Application Level “A” under the GRI. Loh et al. (2017) differentiated the results by industry and status. They considered whether being a Government-Linked Company or a family-owned business should have significant differences. Distinguishing companies with these characteristics is another potential approach. Examining the distinction between different industries is worth considering, given that the TCFD’s recommendations have only singled out several industries for disclosures.

Several elements used by prior studies in a Singapore context will be retained to provide a basis for comparison. The choice of financial metrics will be one of them. This study, however, will seek to propose a robust method of examining sustainability disclosures by examining the text of the sustainability reports.

Methodology

This study is a longitudinal analysis of changes in companies' CSR disclosures for financial periods ending in 2015 to 2021 and their relationship with financial performance.

Research Questions

The research questions are, in the context of Singapore-listed companies:

- RQ 1. What are the most prominent topics in sustainability reports from an exploratory topic modelling analysis? Will this resemble the disclosure themes in frameworks like the GRI?
- RQ 2. How do these topics identified in RQ 1 relate to CFP?
- RQ 3. How do a sustainability report's readability and sentiment score relate to corporate financial performance?
- RQ 4. Are there material differences in the relationship between CSR disclosures and CFP at this point, vis-à-vis prior studies conducted by Tsang, 1998 and Loh et al., 2017?

Data Sources

Some prior studies have opted to use processed data from various sources. For example, Rodgers et al. (2013) used employee, customer satisfaction, and overall social responsibility score from business-ethics.com. Other commonly used sources include ESG metrics from Bloomberg or Thomson Reuters terminals.

This study seeks to use documents provided by the companies when filing their disclosures and use NLP techniques to derive insights directly from the raw text of the sustainability reports. Characteristics of the reports will be derived and discussed in the section on dependent variables. The reports are downloaded either from the company's microsite on sustainability reporting or from the company announcements subsection on the Singapore Exchange.

Financial performance information is taken, extracted from Bloomberg Terminal, and downloaded by ticker. The ratios and standardised financial information, relied on as the independent variable, are derived from annual reports filed with the Singapore Exchange.

Description of Samples

This study will use companies which are component indices of the Straits Times Index (STI). This market capitalisation-weighted index tracks the performance of the top 30 companies listed on SGX.

Companies listed on the Singapore Stock Exchange are of interest, given that they have excelled in tracking and disclosing metrics across historical periods (Loh and Tang, 2021). This aids trend analysis and makes Singapore-listed companies suitable candidates for the analysis. Given also that Singapore is implementing climate reporting on a "comply or explain" basis in their sustainability reports from the

financial year starting in 2022 and is making it mandatory in phases for specific industries in 2023 and 2024 (Reuters, 2021), this is a timely period to take stock of the current state of sustainability reports before new rules come into force.

The samples will consist of sustainability reports of 30 companies in the STI from 2015 to 2021. The start of the time frame is the financial period ending in 2015 when a significant number of companies listed in Singapore began publishing sustainability reports. **Where the sustainability report was unavailable, the "sustainability" or relevant section from the annual report was extracted.** These are denoted in Table 4 as “AR”.

There are a total of 206 reports in the corpus. There were two indices without any sustainability report disclosures. One of them was a new company formed from listing a more extensive group's real estate investment trust - Frasers Logistics and Commercial Trust⁴. Whereas the other company, Venture Corporation Limited, is an exception – which only started to report on sustainability issues when it became mandatory from financial periods ending in 2017. These are denoted in Table 4 as “N/A”.

The majority of companies (19 out of 30, 63%) in the STI maintained a separate sustainability report from 2015 to 2021 ("observation period"). Three companies (10%) continually disclosed sustainability topics in the annual report throughout the observation period. The other companies began disclosing sustainability topics as separate reports during the observation period.

The reports and their source are summarised in Table 4.

⁴ FLCT was listed on SGX Mainboard on 20 June 2016 as Frasers Logistics & Industrial Trust and was subsequently renamed Frasers Logistics & Commercial Trust on 29 April 2020, following their merger with Frasers Commercial Trust.

Information		Sustainability disclosure through							Reference
Company Name	Ticker	2015	2016	2017	2018	2019	2020	2021	Link
COMFORTDELGRO CORPORATION LTD	C52	SUS	SUS	SUS	SUS	SUS	SUS	SUS	https://www.comfortdelgro.com/sustainability
VENTURE CORPORATION LIMITED	V03	N/A	N/A	SUS	SUS	SUS	SUS	SUS	http://venture.listedcompany.com/sustainability_report.html
DBS GROUP HOLDINGS LTD	D05	SUS	SUS	SUS	SUS	SUS	SUS	SUS	https://www.dbs.com/sustainability/reporting/sustainability-report
SINGTEL	Z74	SUS	SUS	SUS	SUS	SUS	SUS	SUS	https://www.singtel.com/about-us/sustainability/sustainability-reports
KEPPEL CORPORATION LIMITED	BN4	SUS	SUS	SUS	SUS	SUS	SUS	SUS	https://www.kepcorp.com/en/sustainability/sustainability-reports/
CAPITALAND INTEGRATED COMM TR	C38U	AR	AR	AR	AR	AR	AR	AR	https://investor.cict.com.sg/ar.html
UNITED OVERSEAS BANK LTD	U11	AR	AR	AR	AR	AR	SUS	SUS	https://www.uobgroup.com/investor-relations/financial/group-annual-reports.html
MAPLETREE LOGISTICS TRUST	M44U	SUS	SUS	SUS	SUS	SUS	SUS	SUS	https://www.mapletreelogisticstrust.com/Investor-Relations/Publications/Annual-Reports.aspx
ASCENDAS REAL ESTATE INV TRUST	A17U	SUS	SUS	SUS	SUS	SUS	SUS	SUS	https://ir.ascendas-reit.com/sustainability_reports.html
OVERSEA-CHINESE BANKING CORP	O39	SUS	AR	SUS	SUS	SUS	SUS	SUS	https://www.ocbc.com/group/sustainability/sustainability-reports.page

YANGZIJANG SHIPBLDG HLDGS LTD	BS6	AR	AR	AR	AR	SUS	SUS	SUS	http://yangzijiang-cn.listedcompany.com/newsroom.html/year/-1
CAPITALAND INVESTMENT LIMITED	9CI	SUS	SUS	SUS	SUS	SUS	SUS	SUS	https://www.capitaland.com/en/about-capitaland/sustainability/sustainability-reports.html
MAPLETREE COMMERCIAL TRUST	N2IU	SUS	SUS	SUS	SUS	SUS	SUS	SUS	https://www.mapletreecommercialtrust.com/Investor-Relations/Publications/Annual-Reports.aspx
THAI BEVERAGE PUBLIC CO LTD	Y92	SUS	SUS	SUS	SUS	SUS	SUS	SUS	https://sustainability.thaibev.com/en/download.php
WILMAR INTERNATIONAL LIMITED	F34	SUS	SUS	SUS	SUS	SUS	SUS	SUS	https://ir-media.wilmar-international.com/sustainability-reports/
MAPLETREE INDUSTRIAL TRUST	ME8U	SUS	SUS	SUS	SUS	SUS	SUS	SUS	https://www.mapletreeindustrialtrust.com/Investor-Relations/Publications/Annual-Reports.aspx
KEPPEL DC REIT	AJBU	SUS	SUS	SUS	SUS	SUS	SUS	SUS	https://www.keppeldcreit.com/en/sustainability/sustainability-report/
CITY DEVELOPMENTS LIMITED	C09	SUS	SUS	SUS	SUS	SUS	SUS	SUS	https://ir.cdl.com.sg/sustainability-reports
SEMBCORP INDUSTRIES LTD	U96	AR	AR	SUS	SUS	SUS	SUS	SUS	https://www.sembcorp.com/en/sustainability/reports-policies
SATS LTD.	S58	AR	AR	SUS	SUS	SUS	SUS	SUS	https://www.sats.com.sg/sustainability/sustainability-reports
UOL GROUP LIMITED	U14	SUS	SUS	SUS	SUS	SUS	SUS	SUS	https://www.uol.com.sg/sustainability/sustainability-reports/
SINGAPORE EXCHANGE LIMITED	S68	AR	AR	AR	AR	AR	AR	AR	https://investorrelations.sgx.com/financial-information/annual-reports

GENTING SINGAPORE LIMITED	G13	SUS	SUS	SUS	SUS	SUS	SUS	SUS	http://www.gentingsingapore.com/#!/en/sustainability/policy/rwscsr
SINGAPORE TECH ENGINEERING LTD	S63	AR	SUS	SUS	SUS	SUS	SUS	SUS	https://www.stengg.com/en/sustainability/sustainability-reports
FRASERS LOGISTICS & COMMERCIAL TRUST	BUOU	N/A	N/A	SUS	SUS	SUS	SUS	SUS	https://www.frasersproperty.com/reits/flct/who-we-are/sustainability#reports
DAIRY FARM INT'L HOLDINGS LTD	D01	AR	AR	AR	AR	AR	AR	AR	https://www.dairyfarmgroup.com/en-US/Investors/Financial-Reports
SINGAPORE AIRLINES LTD	C6L	SUS	SUS	SUS	SUS	SUS	SUS	SUS	https://www.singaporeair.com/en_UK/sg/about-us/information-for-investors/annual-report/
HONGKONG LAND HOLDINGS LIMITED	H78	SUS	SUS	SUS	SUS	SUS	SUS	SUS	https://www.hkland.com/en/sustainability/sustainability-reports
JARDINE CYCLE & CARRIAGE LTD	C07	AR	AR	SUS	SUS	SUS	SUS	SUS	https://www.jcclgroup.com/sustainability
JARDINE MATHESON HLDGS LTD	J36	AR	AR	AR	AR	AR	SUS	SUS	https://ar.jardines.com/2021/Sustainability.html

Table 4: Sustainability disclosures source by company on SGX's STI

Dependent Variables

In research questions RQ 1, RQ 2, and RQ 3, several NLP methods are deployed – tokenisation, Latent Dirichlet Allocation (LDA) for topic modelling, readability assessment, and sentiment analysis.

There are several processing steps. First, the sustainability reports (“corpus”) identified as part of the sample selection will first be tokenised. Before tokenisation, stop words and ancillary words are removed from the tokens. After that, further processing is performed – this could include topic modelling (RQ 1 and RQ 2), readability assessment (RQ 3), and sentiment analysis (RQ 3).

In the further processing steps, the methods would generate an attribute used in the analysis as dependent variables. The derivation of these variables is described in the following subsections.

The code is documented in **Appendix I**.

Tokenisation and Creating Term-Document Matrices

A **token** is a meaningful unit of text used for analysis. Tokenisation, is therefore, the process of splitting the text into tokens – the first step in modelling text data. A graphical example is presented in Figure 12. The method applied uses spaces to split sentences into tokens.

Original Text:

“The reduction of total water consumption and overall water intensity are also key sustainability goals of the Manager.”

As Tokens:

“The”, “reduction”, “of”, “total”, “water”, “consumption”, “and”, “overall”, “water”, “intensity”, “are”, “also”, “key”, “sustainability”, “goals”, “of”, “the”, “Manager”.

(Extracted from Page 41 of Ascendas REIT’s 2019 Sustainability Report)

Figure 12: Illustration of Tokenisation

The Document Term Matrix (DTM) is a mathematical matrix consisting of each document's term (i.e., word) frequency. In this regard, each term corresponds to a 'token', and a document corresponds to each page of the respective sustainability reports. Figure 13 illustrates the DTM using the exact text above.

Document Term Matrix:

$$\begin{bmatrix} D_1 & \text{"of"} & \text{"reduction"} & \text{"total"} & \text{"water"} & \text{"consumption"} & \text{"and"} & \dots & T_t \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ D_n & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix}$$

Whereas n is the number of documents and,

t is the number of terms in the entire corpus of documents, and
the values are representative of the frequency of the terms.

(Extracted from Page 41 of Ascendas REIT’s 2019 Sustainability Report)

Figure 13: Illustration of Creation of Document Term Matrix

The term frequency-inverse document frequency (tf-idf) term weighting is also commonly used in the vector space model, which was used in several studies (Fisher et al., 2016).

This is because tf-idf measures the significance of a word in a corpus or collection of documents. To account for words used more frequently than others, the tf-idf value rises according to the number of times a word appears in the document and is offset by the number of documents in the corpus that contain the term. For example, common words in every document – 'sustainability', 'of', 'and' – will rank lowly. However, if the word 'REIT' often appears in one document but not in the other documents, it is ranked higher and viewed as “relevant”.

The two components of tf-idf are the term frequency and inverse document frequency. For this analysis, a *term* refers to a token (i.e., a word). Whereas the expressions below refer to each term (i.e., word), refer to each document (i.e., each page of the report), and refers to the document set (i.e., the entire corpus of sustainability reports). The tf-idf of a term is the product of these two individual components.

$$tf\ idf(t, d, D) = tf(t, d) * idf(t, D)$$

- **Term frequency:** how frequently each term, t , appears in each document, d . The equation below presents one possible way of adjusting the frequency for exponential effects using a logarithmic function.

$$tf(t, d) = \log (1 + freq(t, d))$$

- **Inverse document frequency:** the logarithm of the total number of documents in a document set, D , divided by the number of documents that contain a term, t (denoted as N).

$$idf(t, D) = \log \left(\frac{N}{count(d \in D: t \in d)} \right)$$

The document-term matrix is the backbone of all the following NLP processing methods applied to the text.

Stop Word Removal, Cleaning, and Lemmatization

Stop words are tokens, or simply words, which are not helpful for the analysis and therefore are removed from the corpus. This is part of cleaning the dataset prior to processing.

With the example presented in Figure 12 featuring a sentence from Ascenda REIT's 2019 Annual Report, it is clear that words such as 'and', 'of', and 'the' are part of a grammatical sentence but do not add any value to the analysis. These words, however, are an attribute (column) in the document-term matrix and keeping them would make computation more costly. Therefore, stop words like these are removed.

Stop word dictionaries from the text mining package for R (Feinerer et al., 2008) are used. English stop words from the Snowball stemmer project and SMART information retrieval system (Lewis et al., 2004) are removed from the corpus.

Similarly, other transformations are performed to clean the text corpus to leave the text corpus with terms (i.e., words) which are meant to be analysed. These transformations include:

1. Removing punctuation symbols,
2. Removing special characters (e.g., '\n'),
3. Converting all words to lowercase – for example, in the case of 'The', and 'the' – the two terms would appear as separate tokens but should be counted as the same. Making the terms all lowercase allows for this to happen.
4. Lemmatise words – the process of grouping together inflected forms of a word to analyse as a single item. For example, 'goal' and 'goals' should be lemmatised to the root word 'goal' to represent one token for this analysis.

Lemmatisation was the preferred approach over stemming as the results are more interpretable. While both are techniques for normalising text to obtain the root form of the word, lemmatisation uses the morphological analysis of words to derive the root form, whereas stemming is a crude heuristic process that chops off the ends of words and removes derivational affixes (Schütze et al., 2008). For example, the word ‘climate’ would be reduced to ‘climat’ by the Porter stemming algorithm (van Rijsbergen et al., 1980) but would be kept as ‘climate’ using Mechura's (2016) English lemmatisation list. As a list is used, not all words are caught. However, a significant number of words are considered; thus, the list fits for purpose.

In addition to the stop word dictionaries provided by Lewis et al. (2004) and Feinerer et al. (2008), some stop words were removed in this analysis. These words are identified through visual inspection of words with the highest frequency. Some examples of words removed are presented in Table 5, whereas the complete list of words is detailed in **Appendix II**.

Words	Category
‘singtel’, ‘wilmar’, ‘cdl’, ‘sia’	Company Names
‘singapore’, ‘thailand’, ‘hong kong’, ‘taiwan’	Country Names
‘www’, ‘com’, ‘non’	Websites and Others

Table 5: Samples of words removed from the document for analysis

As further evidence to support the removal of additional stop words, Figure 14 demonstrates the output of topic modelling prior to removing these stop words. In this example, the number of topics is arbitrarily set as twelve. More on topic modelling using LDA is discussed in the following subsection.

It is observed in Figure 14 that words identified in Table 5 above are part of the top ten (10) words in the topical analysis. However, this is unfortunately not meaningful and does not add value to the analysis of topics when deriving a theme for each topical area. This is a strong impetus for the removal of these other words.

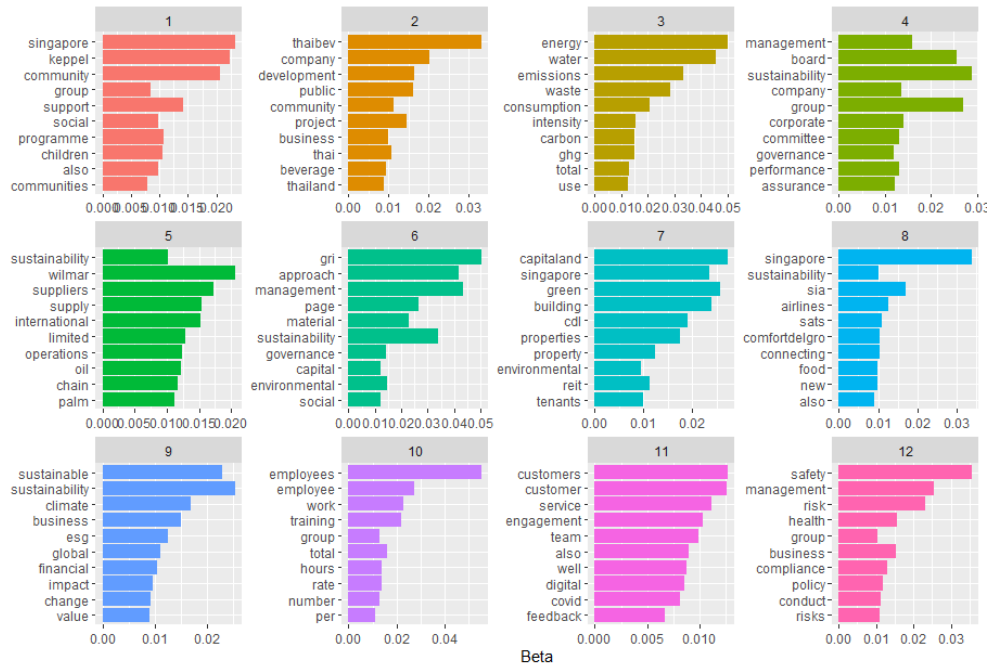


Figure 14: Top ten (10) words per topic, prior to removing stop words

Removing these stop words makes a subsequent performance of tf-idf on the corpus more meaningful. For example, with the definition of tf-idf, words which are identified in Table 5 (e.g. 'singtel') would have been ranked highly for only appearing in Singtel's sustainability reports but not in other companies' reports. The analysis concerns how significant or exceptional sustainability initiatives, metrics, and disclosures affect financial performance rather than individual companies' disclosure.

Latent Dirichlet Allocation (LDA)

Using LDA (Blei et al., 2003), Topic Modelling is a type of dimensionality reduction that employs a probabilistic model to identify co-occurrence patterns of terms corresponding to semantic topics in a set of documents (Crain et al., 2012).

LDA is an improvement from the traditional bag-of-words approach in text feature extraction, where word tokens are mapped to documents. The bag-of-words approach results in huge matrices with most entries as 0. LDA, instead, models relationships between documents and each word token directly, using 'latent variables' as 'bridges.

A Dirichlet distribution over the ‘latent variables (i.e., topics) characterises each document in the corpus and a separate Dirichlet distribution over all tokens’ characteristics for each topic. Figure 15 illustrates this.

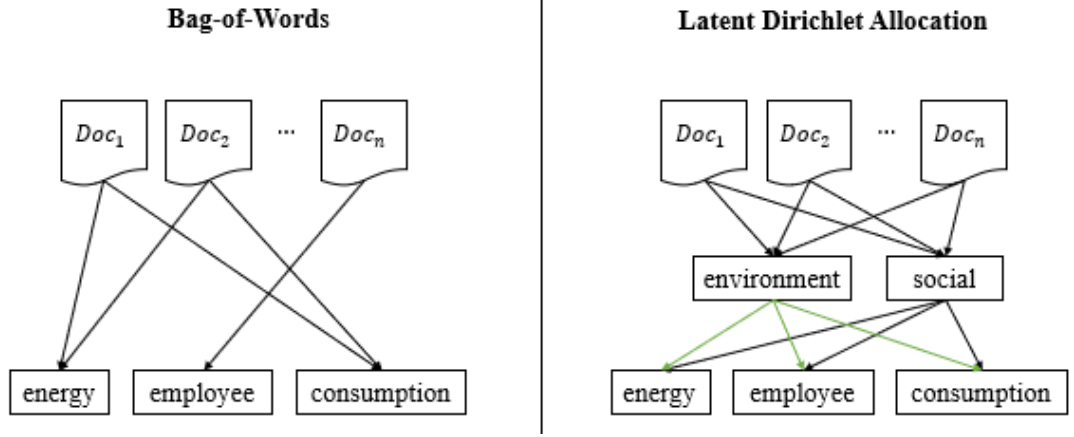


Figure 15: Differences between Bag-of-Words and Latent Dirichlet Allocation in text feature extraction

For each document, a vector of length k is obtained, where k represents the number of pre-defined topics, T . Therefore, appending the vectors together for each document, d , would give an n by k matrix to be used for the analysis, whereas p represents the probability distribution of the document over k topics.

$$\begin{bmatrix} D_1 & T_1: Env. & T_2: Social & \cdots & T_k \\ p_{1,1} & p_{1,2} & \cdots & p_{1,k} \\ \vdots & & & \vdots \\ D_n & p_{n,1} & p_{n,2} & \cdots & p_{n,k} \end{bmatrix}$$

Selecting the Optimal Number of Topics

A key parameter of LDA is the pre-defined number of topics, k . This is evaluated with LDA model selection metrics. In this analysis, the ‘CaoJuan2009’ (Cao et al., 2009) and ‘Deveaud2014’ (Deveaud et al., 2014). A given set’s optimal number of topics is reached when the overall dissimilarity between topics achieves its maximum. ‘CaoJuan2009’ selects the best LDA model based on density, whereas ‘Deveaud2014’ forces on maximising conceptual coherence. As part of this analysis, the themes should

be interpretable. The balance is between enough topics to identify key themes for separate analysis but not too many that the themes are not interpretable. For example, the method `FindTopicsNumber()` from the R package `ldatuning` is used to tune the parameter – number of topics.

The results of metrics for “CaoJuan2009” and “Deveaud2014” are presented in Figure 16. A minimal value for “CaoJuan2009” is more optimal, whereas a maximal value for “Deveaud2009” is more optimal. The metrics corresponding to each parameter of the number of topics and each metric is also presented in **Appendix III**.

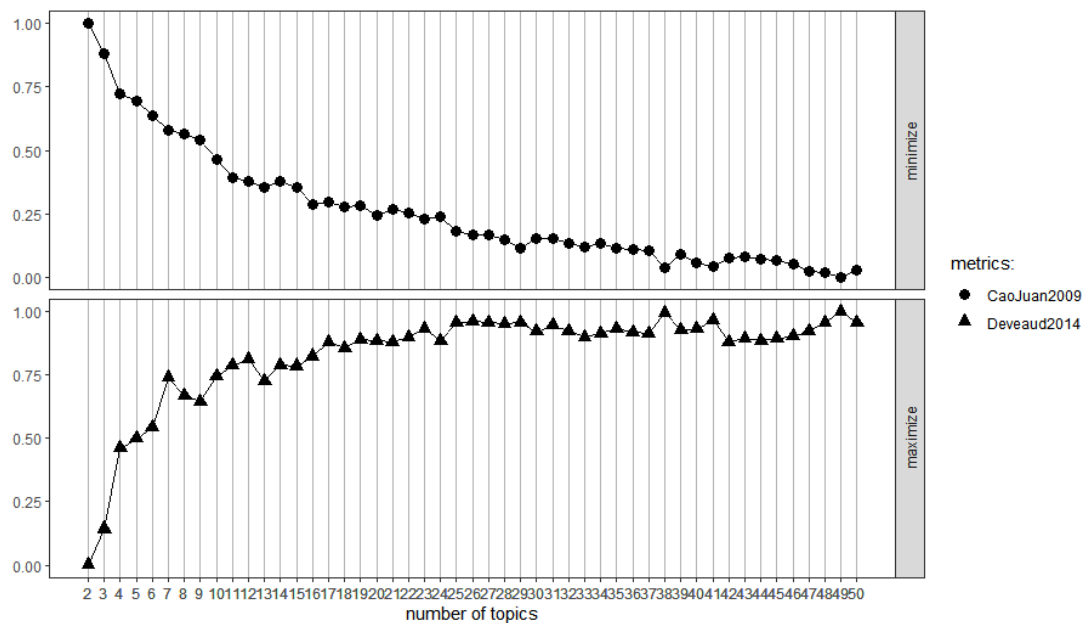


Figure 16: *CaoJuan2009* (Cao et al., 2009) and *Deveaud2014* (Deveaud et al., 2014) metric scores by the parameter – number of topics

The “Deveaud2014” metric begins to hover between 3.01 – 3.06 when the number of topics increases beyond 17. Although the “CaoJuan2009” metric still decreases, the decreases after topics are marginal compared to the initial values. In the interest of having more interpretable topics, the lowest parameter where an increase in the number of topics brings a significant improvement of the metric is seven (7) topics. This shall be picked for the subsequent analysis.

Determining Topics

To determine the clusters of topics, Gibbs sampling is used. Gibbs sampling is an algorithm that is based off Monte Carlo Markov Chains. The goal is to find the conditional probability distribution of a token's ("word") topical assignment conditioned on other topic assignments. For a word, w , in a document, d , in a topic k , expressed mathematically:

$$p(z_{d,n} = k | \vec{z}_{d,n}, \vec{w}, \alpha, \lambda) = \frac{n_{d,k} + \alpha_k}{\sum_i^k n_{d,i} + \alpha_i} * \frac{v_{(k,w)} + \lambda_{w,d,n}}{\sum_i v_{k,i} + \lambda_i}$$

Where:

- $n_{d,k}$ is the number of times document d uses topic k
- $v_{k,w}$ is the number of times topic k uses the word w
- α_k is the Dirichlet parameter for the document to topic distribution
- λ_w is the Dirichlet parameter for topic to word distribution

The tokens start with a random topic assignment. After that, the topics are reassigned based on the highest conditional probabilities. In this analysis, the number of iterations is set at 500. This algorithm may be repeated indefinitely until the conditional probabilities converge to a maximal point. However, this was not proceeded with as 500 iterations produced an intuitive and interpretable result with sufficient dissimilarity.

The results from this LDA modelling are presented in the subsequent subsection as part of the discussion and results.

Text Readability

The originally Flesch Reading Ease Score will be used for its simplicity. However, other factors were not included in the mix to avoid autocorrelation in the regression analysis. As part of these metrics, lexical density is assessed. This represents the number of content words (e.g. ‘sustainability’ or ‘company’) vis-à-vis the number of grammatical words (e.g. ‘if’, ‘but’, or ‘will’). A higher lexical density is interpreted as a higher textual complexity due to a higher conceptual load (Halliday, 1989; Harrison and Bakker, 1998).

Flesch (1948) defines the Flesch Reading Ease Score as:

$$RE = 206.835 - (1.015 * ASL) - (84.6 * ASW)$$

Whereas:

RE is the Readability Ease, where $(0 \leq RE \leq 100)$

ASL is the Average Sentence Length

ASW is the Average Syllables per Word

Whereas the ranges of scores may be interpreted as illustrated in Table 6:

Readability Ease	Category
90-100	Very Easy
80-89	Easy
70-79	Fairly Easy
60-69	Standard
50-59	Fairly Difficult
30-49	Difficult
0-29	Very Confusing

*Table 6: Interpretation of Readability Ease from the Flesch Reading Ease Score
(Flesch, 1948)*

Each sustainability report will be given a Readability Ease score. This uses the `readability()`⁵ method, with the Flesch Reading Ease Score of the R package `sylcount`⁶. This shall be used as a dependent variable in the subsequent analysis to answer RQ 3.

Text Sentiment

Sentiment analysis uses the sentiment content of individual words to represent the sentiment content of the whole text. This is done based on pre-defined lists of words and scores, known as lexicons. These lexicons are based on single words and contain many English words, whereas the words are assigned scores for positive or negative sentiment. After tokenisation, each ‘token’ in the sustainability report is one word. This word is joined with the lexicons to derive a score.

The lexicons used are decisive of the result. Words in different contexts mean different things. Taking the below extract from Ascendas REIT’s 2019 Sustainability Report:

Original Text:

“The reduction of total water consumption and overall water intensity are also key sustainability goals of the Manager.”

(Extracted from Page 41 of Ascendas REIT’s 2019 Sustainability Report)

The word ‘reduction’ may carry a negative sentiment in general terms, although in the context of sustainability text, it may be either neutral or positive. Another example is the word ‘risk’. In a general context, it may be negative. However, in a finance or sustainability context, ‘risk’ is an object rather than a descriptor.

⁵ Documentation is available through [readability {koRpus}](#).

⁶ Documentation is available at [Package ‘sylcount’](#).

Rutherford (2003) observed the “obfuscation hypothesis”, which posits that companies will make unfavourable news more difficult to decode by using positive language, especially in times of crisis. Therefore, this is an element of assessment in RQ 3. To assess the sentiment of tone used in the sustainability report, several general and one finance-specific lexicon will be used. The lexicons used are summarised in Table 7.

Purpose	Lexicon	Reference	Description	Score, x for each word
General	afinn	(Nielsen, 2011)	Polarity and intensity of words (2,476 words)	$-5 \leq x \leq 5$ (discrete)
General	nrc	(Mohammad and Turney, 2013)	Polarity of words (5,468 words)	$x: \{-1, 1\}$
General	bing	(Hu and Liu, 2004)	Polarity of words (6,874 words)	$x: \{-1, 1\}$
General	jockers	(Jockers, 2017)	Incorporates emotional shifts in text – polarity and intensity (10,738 words)	$-1 \leq x \leq 1$ (continuous)
Finance	loughran	(Loughran and McDonald, 2011)	Assigns several categories – ‘positive’ and ‘negative’ will be converted respectively.	N/A

Table 7: Lexicons used in this Analysis

For this analysis, each sustainability report will be given a sentiment score, an average of the sentiment scores from using several lexicons. This is done to account for different words present in the lexicon that may contribute to different scores. The `get_sentiments()` method from the R package **tidytext** is used. Other lexicons are

taken from the R packages **syuzhet** and **lexicon**. The sentiment score shall also be used as a dependent variable in the subsequent analysis to answer RQ 3.

Notably, qualifiers before words are not accounted for using these mechanics. For example, the phrase ‘not satisfied’ will be analysed as two separate words – ‘not’ and ‘satisfied’. At the same time, ‘satisfied’ is typically interpreted as a positive sentiment. Nonetheless, given the volume of text in the sustainability report, it is expected that this effect is minimal, and together using single words (“unigrams”) will give an indicative score of the report’s sentiment for this analysis.

Samples of output from the sentiment analysis are shown below, with the first few words from Ascendas REIT’s 2015 Sustainability Report used as an example. The outputs are summarised in Table 8.

Lexicon	Sample Output
AFINN	<pre>> text_tokenised %>% + inner_join(get_sentiments("afinn")) Joining, by = "word" # A tibble: 211,189 x 5 page Company Year word value <int> <chr> <int> <chr> <dbl> 1 1 1 ASCENDAS REAL ESTATE INV TRUST 2015 reach 1 2 2 2 ASCENDAS REAL ESTATE INV TRUST 2015 responsible 2 3 3 2 ASCENDAS REAL ESTATE INV TRUST 2015 responsible 2 4 4 2 ASCENDAS REAL ESTATE INV TRUST 2015 solutions 1 5 5 2 ASCENDAS REAL ESTATE INV TRUST 2015 safety 1 6 6 2 ASCENDAS REAL ESTATE INV TRUST 2015 safety 1 7 7 3 ASCENDAS REAL ESTATE INV TRUST 2015 pleased 3 8 8 3 ASCENDAS REAL ESTATE INV TRUST 2015 success 2 9 9 3 ASCENDAS REAL ESTATE INV TRUST 2015 proud 2 10 10 3 ASCENDAS REAL ESTATE INV TRUST 2015 prepared 1 # ... with 211,179 more rows</pre>
NRC	<pre>> text_tokenised %>% + inner_join(get_sentiments("nrc")) Joining, by = "word" # A tibble: 916,718 x 5 page Company Year word sentiment <int> <chr> <int> <chr> <chr> 1 2 2 ASCENDAS REAL ESTATE INV TRUST 2015 responsible positive 2 2 2 ASCENDAS REAL ESTATE INV TRUST 2015 responsible trust 3 2 2 ASCENDAS REAL ESTATE INV TRUST 2015 talent positive 4 2 2 ASCENDAS REAL ESTATE INV TRUST 2015 management positive 5 2 2 ASCENDAS REAL ESTATE INV TRUST 2015 management trust 6 2 2 ASCENDAS REAL ESTATE INV TRUST 2015 assessment surprise 7 2 2 ASCENDAS REAL ESTATE INV TRUST 2015 assessment trust 8 2 2 ASCENDAS REAL ESTATE INV TRUST 2015 accountability positive 9 2 2 ASCENDAS REAL ESTATE INV TRUST 2015 accountability trust 10 2 2 ASCENDAS REAL ESTATE INV TRUST 2015 responsible positive # ... with 916,708 more rows</pre>

Bing	<pre>> text_tokenised %>% + inner_join(get_sentiments("bing")) Joining, by = "word" # A tibble: 237,948 x 5 page Company Year word sentiment <int> <chr> <int> <chr> <chr> 1 1 ASCENDAS REAL ESTATE INV TRUST 2015 sustainability positive 2 2 ASCENDAS REAL ESTATE INV TRUST 2015 sustainability positive 3 2 ASCENDAS REAL ESTATE INV TRUST 2015 well positive 4 2 ASCENDAS REAL ESTATE INV TRUST 2015 talent positive 5 2 ASCENDAS REAL ESTATE INV TRUST 2015 well positive 6 3 ASCENDAS REAL ESTATE INV TRUST 2015 sustainability positive 7 3 ASCENDAS REAL ESTATE INV TRUST 2015 pleased positive 8 3 ASCENDAS REAL ESTATE INV TRUST 2015 success positive 9 3 ASCENDAS REAL ESTATE INV TRUST 2015 sustainability positive 10 3 ASCENDAS REAL ESTATE INV TRUST 2015 proud positive # ... with 237,938 more rows</pre>
Jockers	<pre>> text_tokenised %>% + inner_join(key_sentiment_jockers) Joining, by = "word" # A tibble: 573,182 x 5 page Company Year word value <int> <chr> <int> <chr> <dbl> 1 1 ASCENDAS REAL ESTATE INV TRUST 2015 deepening -0.25 2 1 ASCENDAS REAL ESTATE INV TRUST 2015 sustainability 1 3 2 ASCENDAS REAL ESTATE INV TRUST 2015 sustainability 1 4 2 ASCENDAS REAL ESTATE INV TRUST 2015 responsible 0.25 5 2 ASCENDAS REAL ESTATE INV TRUST 2015 well 0.8 6 2 ASCENDAS REAL ESTATE INV TRUST 2015 talent 0.5 7 2 ASCENDAS REAL ESTATE INV TRUST 2015 management 0.4 8 2 ASCENDAS REAL ESTATE INV TRUST 2015 accountability 0.8 9 2 ASCENDAS REAL ESTATE INV TRUST 2015 responsible 0.25 10 2 ASCENDAS REAL ESTATE INV TRUST 2015 management 0.4 # ... with 573,172 more rows</pre>
Loughran	<pre>> text_tokenised %>% + inner_join(get_sentiments("loughran")) Joining, by = "word" # A tibble: 162,194 x 5 page Company Year word sentiment <int> <chr> <int> <chr> <chr> 1 2 ASCENDAS REAL ESTATE INV TRUST 2015 satisfaction positive 2 3 ASCENDAS REAL ESTATE INV TRUST 2015 pleased positive 3 3 ASCENDAS REAL ESTATE INV TRUST 2015 success positive 4 3 ASCENDAS REAL ESTATE INV TRUST 2015 challenges negative 5 3 ASCENDAS REAL ESTATE INV TRUST 2015 comply constraining 6 3 ASCENDAS REAL ESTATE INV TRUST 2015 regulation litigious 7 3 ASCENDAS REAL ESTATE INV TRUST 2015 enabled positive 8 3 ASCENDAS REAL ESTATE INV TRUST 2015 stringent negative 9 3 ASCENDAS REAL ESTATE INV TRUST 2015 committed constraining 10 3 ASCENDAS REAL ESTATE INV TRUST 2015 committed constraining # ... with 162,184 more rows</pre>

Table 8: Sample Output of Word Sentiments by Lexicon

The scores are assigned to individual words which were present in the documents. The sentiment score of the document is a summation of scores from all found words in the document. Where the sentiment returned is a category, as in NRC, Bing, and Loughran, only ‘positive’ and ‘negative’ sentiments are kept. The sentiment score for these lexicons is the number of positive words less the number of negative words.

Given that each lexicon has a different set of words, the arithmetic mean of all sentiment scores from the five lexicons used is representative of the document's sentiment score.

As the sentiment scores are a summation of the scores from individual words across the whole document, simply taking the summation may be misleading as a report with more words may have a higher sentiment score with more positive words. Therefore, the average summation of sentiment scores divided by the number of pages is used for this analysis. An illustration is presented below, using Dairy Farm International (“DFI”) Holding’s 2016 Sustainability Report.

Step 1: The average of five lexicons are calculated

Coy.	Year	Page	AFINN	Bing	NRC	Jockers	Loughran	Average
DFI	2016	1	16	11	32	0	22.05	16.21
DFI	2016	2	18	11	18	7	26.95	20.19

Step 2: The average of the average sentiment scores, S , for document d is calculated

$$S_d = \frac{16.21 + 20.19}{2} = 18.2$$

This is the sentiment score that is used for further analysis.

Independent Variables

This analysis seeks to determine the relationship between sustainability disclosures and the corresponding financial performance of the company.

In a previous analysis of Singapore-listed companies, Wahyuningrum et al. (2021) used Return on Assets (ROA) as an indicator of profitability, Loh et al. (2017) used the Market Value (MV) four months after the disclosure date, and Khaveh et al. (2012) used the Revenue and Share Price as a proxy for financial performance. These studies use metrics from the same years of disclosures. To provide a basis for comparison, as suggested in RQ 4, similar measures will be adopted – the ROA and the MV of the company four months after the disclosure date. In an analysis of Brazilian companies, ROA was also used by Aggarwal (2013) for comparison against Indian companies and Ching et al. (2017).

Whereas ROA and MV are defined respectively:

$$ROA = \frac{Net\ Income}{Total\ Assets}$$

$$MV = Share\ Price * (\#\ Total\ Outstanding\ Shares)$$

In addition, Rodgers et al. (2017) assessed financial viability with the Zmijewski score. This was formulated in Zmijewski (1984) by regressing various factors over 40 bankrupt and 800 companies, which were going concerns to determine the likelihood of bankruptcy. In discussions about sustainability, this study will also analyse the relationship between sustainability disclosures and the firm's financial viability, a subset of financial performance to be a going concern.

The Zmijewski Score (Zmijewski, 1984), is defined as:

Zmijewski score

$$\begin{aligned} &= -4.336 - 4.513 * \frac{\text{Net income}}{\text{Total assets}} + 5.679 * \frac{\text{Total liabilities}}{\text{Total assets}} \\ &+ 0.004 * \frac{\text{Current assets}}{\text{Current liabilities}} \end{aligned}$$

The Zmijewski Score considers the profitability (return on assets), solvency (liabilities to assets), and liquidity (current ratio) of the company. The output is a continuous variable, whereas higher (more positive) scores indicate that the companies are likely to go bankrupt (i.e., less financially viable; worse financial performance) and lower (more negative scores) indicate that the companies are less likely to go bankrupt (i.e., more financially viable; better financial performance).

The ROA, MV, and Zmijewski scores will be used for this analysis. Each of these independent variables will analyse by dependent variables. This is explained in the following subsection on the models and approach.

Models and Approach

The variables of interest outlined in the previous section are summarised in Table 9.

Dependent Variables <i>(Sustainability Disclosures)</i>	Independent Variables <i>(Financial Performance)</i>
Topic Composition (TC) (RQ 2)	Return on Assets (ROA)
Flesch Reading Ease Score (RE) (RQ 3)	Market Value (MV)
Weighted-Sentiment Scores (S) (RQ 3)	Zmijewski Score (ZJS)

Table 9: Variables used in the analysis.

These variables will be iterated in different combinations in response to RQ 2, RQ 3, and, subsequently, RQ 4. In addition, the Flesch Reading Ease Score, *RE*, and Weighted-Sentiment Scores, *S*, will be analysed together. This is done considering the signalling effects and obfuscation hypothesis discussed in Rutherford (2003).

The relationships are expressed mathematically in the subsequent sections. The generic hypotheses for each of the following six iterations are:

For overall model significance (F-test):

$$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_1: 1 \text{ or more parameters } \neq 0$$

For individual variable significance:

$$H_0: \beta_p = 0$$

$$H_1: \beta_p \neq 0$$

Whereas β_p represents the regression coefficients, and the significance level is 5%.

Analysis on Return on Assets (ROA)

For the analysis of ROA, an ordinary least squares multiple linear regression is performed. This is used in Wahyunigram et al. (2021). Whilst Rodgers et al. (2007) and Rodgers et al. (2013) have identified benefits of partial least squares in addressing multicollinearity, in this case, the variables were intentionally selected to avoid autocorrelation. In favour of Occam's razor, the ordinary least squares regression is favoured.

For RQ 2:

$$ROA_t = \beta_0 + \beta_1 TC_{t,2} + \beta_2 TC_{t,2} + \dots + \beta_k TC_{t,k} + \epsilon$$

Where k is the number of topics, and t is the time period.

For RQ 3:

$$ROA_t = \beta_0 + RE_t + S_t + \epsilon$$

Analysis on Market Value (MV)

The Ohlson's model (Ohlson, 1995) will be used as a baseline model. The Ohlson's model was also used in Loh et al. (2017). For comparability to prior studies, this is preferred – to see if the sustainability indications have an additional implication. Similarly, ordinary least squares regression is used. In this case, for the individual variable significance, the focus is on the coefficients that relate to the dependent variables highlighted – i.e., β_4 , β_5 , etc.

For RQ 2:

$$MV_{t+4} = \beta_0 + \beta_1 BV_t + \beta_2 EARN_t + \beta_3 EARN_t * NEG_t \\ + (\beta_4 TC_1 + \beta_5 TC_2 + \dots \beta_{k+3} TC_3) + \epsilon$$

For RQ 3:

$$MV_{t+4} = \beta_0 + B_1BV_t + \beta_2EARN_t + \beta_3EARN_t * NEG_t + (\beta_4RE_t + \beta_5S_t) + \epsilon$$

Where:

MV_{t+4} is the market value four months after financial year-end of the company.

BV_t is the book value of common equity and the year-end of the company.

$EARN_t$ is the earnings before extraordinary items for the financial year.

NEG_t is a dummy variable equal to 1 if earnings are negative, and 0 otherwise.

Analysis on Zmijewski Score (ZJS)

Like the analysis on ROA, this analysis will use an ordinary least squares multiple linear regression analysis.

For RQ 2:

$$ZJS_t = \beta_0 + \beta_1TC_{t,2} + B_2TC_{t,2} + \dots + B_kTC_{t,k} + \epsilon$$

Where k is the number of topics, and t is the time period.

For RQ 3:

$$ZJS_t = \beta_0 + RE_t + S_t + \epsilon$$

Discussion and Results

Data Exploration

In this section, summary statistics about the sustainability reports are presented. In addition, there will first be a very high-level overview of the sustainability reports in this sample space and a presentation about the words in the sustainability reports.

Number of Pages and Words of Sustainability Reports

Although quantity does not necessarily reflect quality, one way to gauge the extent of disclosures is to observe the changes in the number of pages. The results are summarised in Table 10, Table 11 **and** Figure 17.

The average number of pages in sustainability reports has risen over the years. For example, the Singapore Exchange had mandated sustainability reporting on a “comply or explain” basis from the financial year ending in 2017, which saw a significant increase in the average number of pages by 34% between the financial years ending in 2016 and 2017. The increases, however, were more significant in recent years. This could be owed to improved processes and an increased ability to make such disclosures.

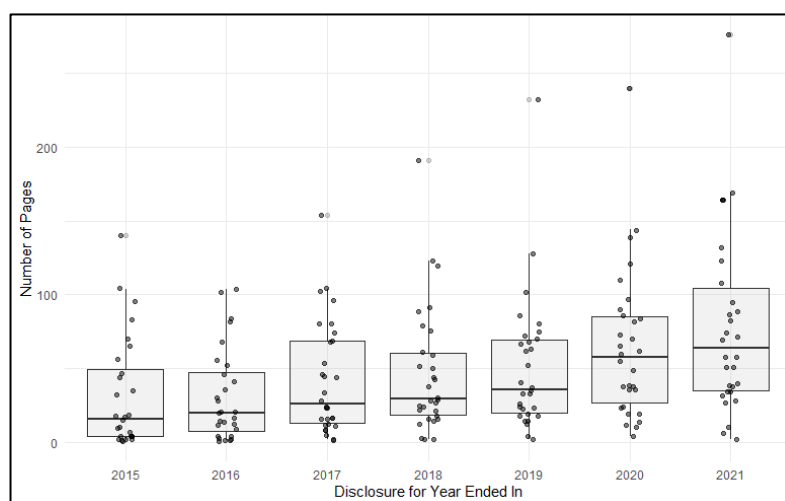


Figure 17: Number of Pages of Sustainability Reports by Financial Year

Year	Count of Reports	Sum of Pages	Average Pages	Min. Pages	Max. Pages	Std. Dev. Pages
2015	27	830	30.74	1	140	37.70
2016	27	814	30.15	1	104	31.10
2017	29	1,204	41.52	2	154	38.43
2018	29	1,327	45.76	2	191	42.87
2019	29	1,461	50.38	2	232	46.69
2020	29	1,852	63.86	4	240	51.39
2021	30	2,373	79.10	2	276	61.52

Table 10: Summary Statistics of Number of Pages in the Sustainability Report by Financial Year

	Count of Reports	Sum of Pages	Average Pages	Min. Pages	Max. Pages	Std. Dev. Pages
Total	200	9,861	49.31	1	276	47.75

Table 11: Summary Statistics of Number of Pages in the Sustainability Report – Overall

Whereas the minimum and maximum number of pages is listed in Table 12. The disclosure trends are consistent throughout the observation period of financial years ending from 2015 to 2021. Dairy Farm International Holdings Limited (“DFI”), Jardine Matheson Holdings Limited, and Jardine Cycle and Carriage Limited (“JCC”) had the shortest sustainability reports. DFI and JCC are subsidiaries of the Jardine Matheson Group. Thai Beverage Public Company Limited and City Developments Limited presented the maximum number of pages.

Year	Minimum of Pages	Maximum of Pages
2015	DAIRY FARM INT'L HOLDINGS LTD JARDINE CYCLE & CARRIAGE LTD	THAI BEVERAGE PUBLIC CO LTD
2016	JARDINE CYCLE & CARRIAGE LTD SATS LTD.	CITY DEVELOPMENTS LIMITED
2017	DAIRY FARM INT'L HOLDINGS LTD JARDINE MATHESON HLDGS LTD	THAI BEVERAGE PUBLIC CO LTD

2018	DAIRY FARM INT'L HOLDINGS LTD JARDINE MATHESON HLDGS LTD	THAI BEVERAGE PUBLIC CO LTD
2019	JARDINE MATHESON HLDGS LTD	THAI BEVERAGE PUBLIC CO LTD
2020	DAIRY FARM INT'L HOLDINGS LTD	THAI BEVERAGE PUBLIC CO LTD
2021	DAIRY FARM INT'L HOLDINGS LTD	THAI BEVERAGE PUBLIC CO LTD

Table 12: Companies with the Minimum and Maximum Number of Pages in their Sustainability Reports by Financial Year

The number words in the sustainability reports can be obtained by way of tokenisation – where each word in the document becomes one attribute. Figure 18 below presents these observations visually by year.

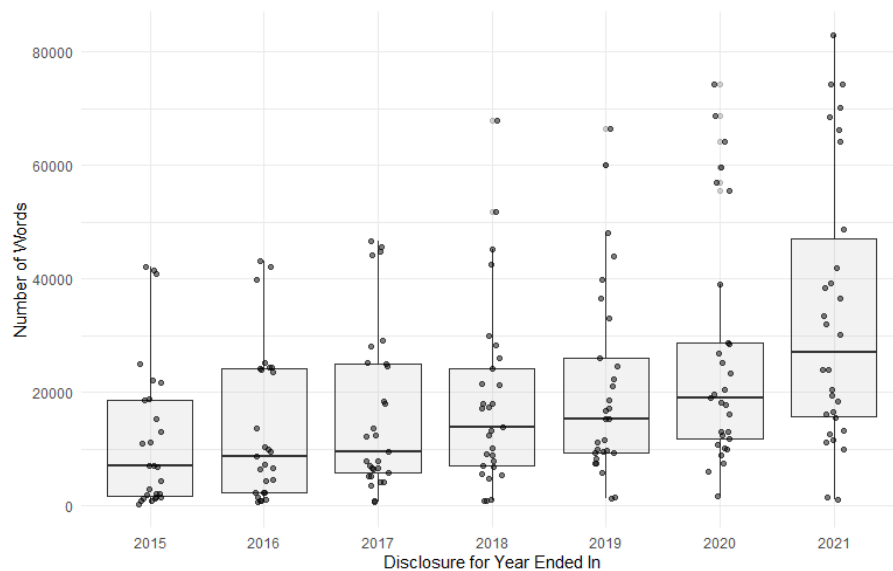


Figure 18: Number of Words in Sustainability Reports by Financial Year

Like the number of pages, there was an increase in the number of words each year. Albeit the increases in the average number of words in 2020 and 2021 appear more substantial vis-à-vis the number of pages. Table 13 contains the summary statistics of the number of words in sustainability reports by year, Table 14 contains the overall summary statistics, and Table 15 features the top and bottom companies.

Year	Count of Reports	Sum of Words	Average Words	Min. Words	Max. Words	Std. Dev. Words
2015	27	323,253	11,972	280	42,088	13,053
2016	27	364,625	13,505	773	43,207	13,314
2017	29	469,719	16,197	652	46,647	14,388
2018	29	538,091	18,555	790	68,001	16,204
2019	29	608,277	20,975	1,360	66,461	16,992
2020	29	767,695	26,472	1,810	74,222	20,872
2021	30	1,016,635	33,888	1,185	82,926	24,121

Table 13: Summary Statistics of Number of Pages in the Sustainability Report by Financial Year

	Count of Reports	Sum of Words	Average Words	Min. Words	Max. Words	Std. Dev. Words
Total	200	4,088,295	20,441	280	82,926	18,713

Table 14: Summary Statistics of Number of Words in the Sustainability Report – Overall

Companies which had the least disclosure in number of pages were largely consistent with the number of words. However, there were variances in the greatest number of words and pages for 2015-2018. While Thai Beverage Public Co Ltd had the most number of pages consistently, it appears that City Developments and Keppel Corporation Limited delivered more textual content than pages. Suggesting that Thai Beverage Public Co Ltd's initially reports were relatively more graphical.

Year	Minimum of Words	Maximum of Words
2015	DAIRY FARM INT'L HOLDINGS LTD	CITY DEVELOPMENTS LIMITED
2016	SATS LTD.	KEPPEL CORPORATION LIMITED
2017	DAIRY FARM INT'L HOLDINGS LTD	KEPPEL CORPORATION LIMITED
2018	DAIRY FARM INT'L HOLDINGS LTD	CITY DEVELOPMENTS LIMITED
2019	JARDINE MATHESON HLDGS LTD	THAI BEVERAGE PUBLIC CO LTD

2020	DAIRY FARM INT'L HOLDINGS LTD	THAI BEVERAGE PUBLIC CO LTD
2021	DAIRY FARM INT'L HOLDINGS LTD	THAI BEVERAGE PUBLIC CO LTD

Table 15: Companies with the Minimum and Maximum Number of Words in their Sustainability Reports by Financial Year

Joining the information on the number of words and pages can present some insights. Naturally, there is a positive relationship between the number of pages and number of words. Figure 19 shows the number of pages versus the number of words.



Figure 19: Number of words versus number of pages in sustainability reports, with a trend line

The scatterplot demonstrates that the distribution of the number of words and number of pages in sustainability reports is right-skewed. The majority of reports tend to have a lower (<15k words, <35 pages) number of pages, with the average being inflated by

companies like Thai Beverages, City Developments, and Keppel Corporation, which have an exceptionally high extent of disclosure. On average, there are 388 words per page across sustainability reports in the sample group.

Reports on either side of the trend line allow for inference on the composition of the reports – whether the reports have more textual content per page or more visual content per page. For example, the disclosure style is generally consistent among the most voluminous disclosures throughout the years. Thai Beverage consistently has a lower word-to-page ratio, indicating a relatively higher proportion of visual content. Whereas Keppel Corporation and City Developments have a higher word-to-page ratio. Hrasky (2012), in relation to the obfuscation hypothesis, suggested that attractive imagery for window-dressing and greenwashing is a possible disclosure strategy. While greenwashing with visual disclosures is not the focus of this analysis, it is noteworthy to keep organisations with a greater extent of visual disclosures in mind when furthering the analysis.

Key Topics in Sustainability Reports

RQ 1. What are the most prominent topics in sustainability reports from an exploratory topic modelling analysis? Will this resemble the disclosure themes in frameworks like the GRI?

Using the Gibbs Sampling method for Latent Dirichlet Allocation highlights the common clusters of word-to-document distributions; it allows for an analysis of the key themes across sustainability reports in the sample. The preceding subsection determined that using the parameter of seven (7) topics was the optimal trade-off between performance and interpretability. The top ten words per topic are presented in Figure 20. The proposed interpretation of these topics is suggested below in Table 16.

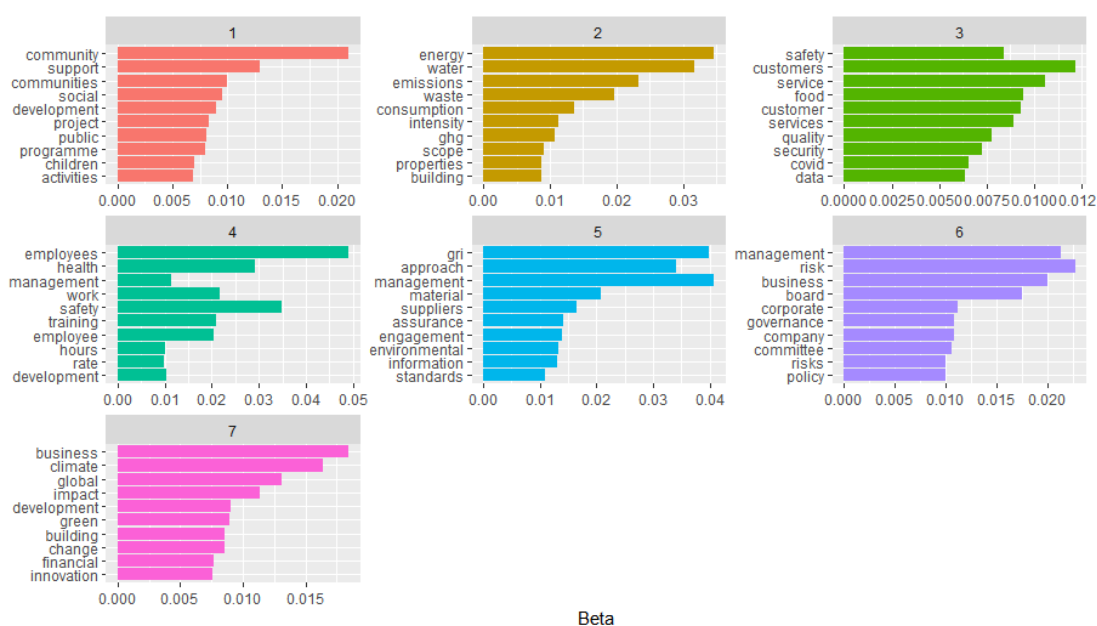


Figure 20: Top ten words by identified topic

Topic	Theme	Words
1	Social / Community	Community, Social, Support, Development, Children, Activities
2	Environment / Resources	Energy, Emissions, Waste, Water, Consumption, GHG
3	Social / Customers	Customer, Safety, Service, Security, Quality
4	Social / Employees	Employees, Work, Safety, Development, Training
5	Methodology	GRI, Material, Assurance, Information, Standards
6	Governance	Board, Corporate, Governance, Committee, Risk, Policy
7	Environment / Climate	Climate, Impact, Green, Building

Table 16: Proposed Interpretation of Topics

Alignment with Disclosure Themes Required in GRI

GRI 3 specifies that the material topics are topics that represent an organisation's most significant impacts on the economy, environment, and people, including impacts on their human rights. Three of the identified optimal number of seven topics are related to people. There were also topics for the environment – resources and climate.

Although, specific references to human rights are not present. This could be embedded within the existing topics. There are some broad themes similar to the classification proposed by Kang and Kim (2022) in Table 1.

GRI 2 contains disclosure guidelines for organisations to provide information about their reporting practices, activities, and workers. One of the seven themes identified pertains to methodology and disclosures.

Given that the GRI is adopted in 99% of listed companies in Singapore (GRI and NUS, 2022), it is within expectations that the themes identified from topic modelling are influenced by the disclosure themes required by the GRI.

Distribution of Topics in Sustainability Reports

Overall, the distribution of topics across all reports is consistent across the years. For example, it appears that there is around 40% of the content dedicated to disclosures on the social dimension – community, customers, and employees for Singaporean companies in the disclosure sample space. This is shown in Figure 21.

Because of climate disclosure requirements implemented by SGX, the proportion of climate-related disclosures has increased over the years. This is consistent with our expectations. As a result, the disclosures for community-related initiatives and governance topics have fallen in relativity.

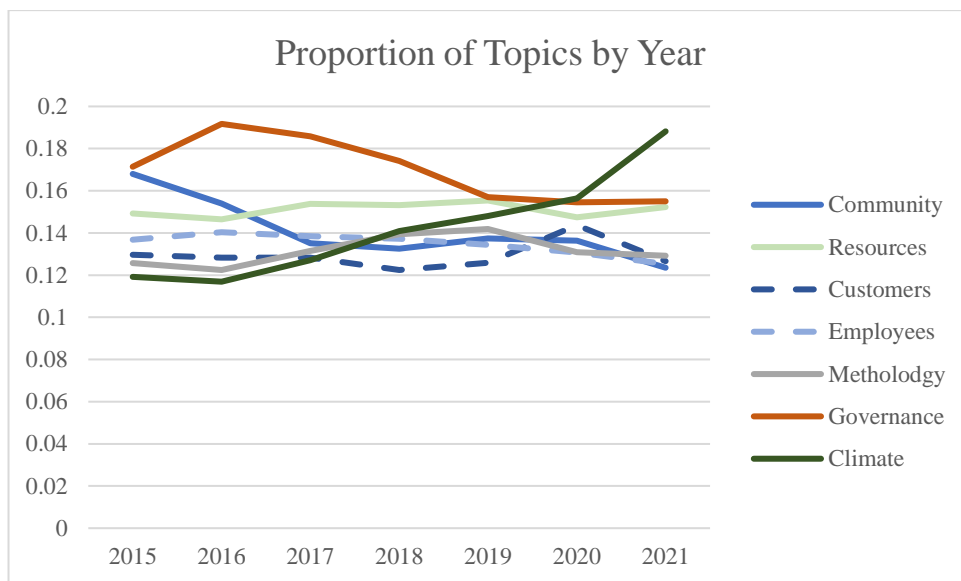
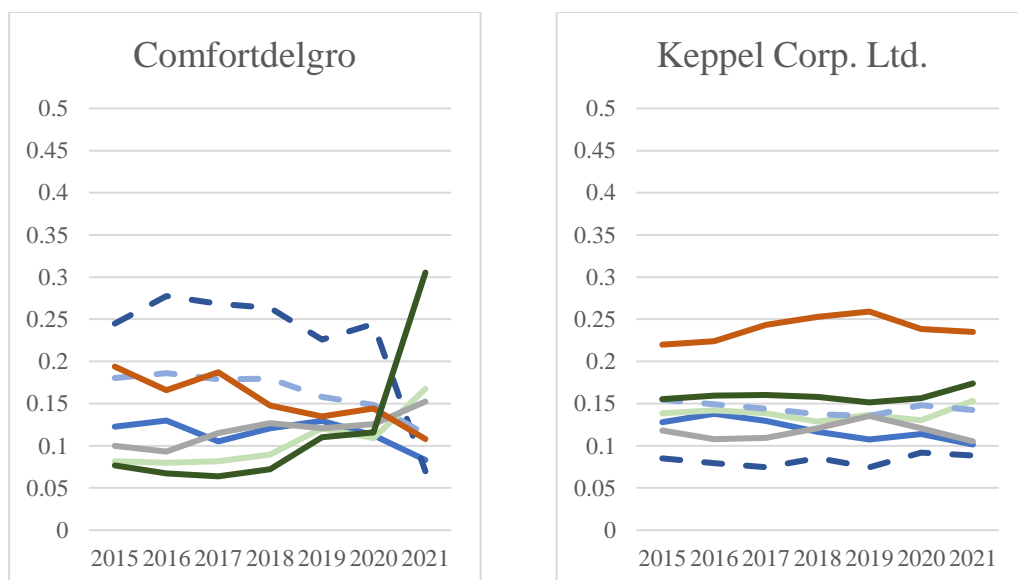


Figure 21: Distribution of Topics across the Years

The distribution of topics is consistent within companies throughout the observation period of 2015 to 2021. However, there is a difference in the proportion each topic takes up by different companies. Contrasting examples are featured using: (i) Comfortdelgro Corporation Limited, (ii) Keppel Corporation Limited, (iii) Sembcorp Industries Ltd, and (iv) Wilmar International Limited. This is shown in Figure 22.



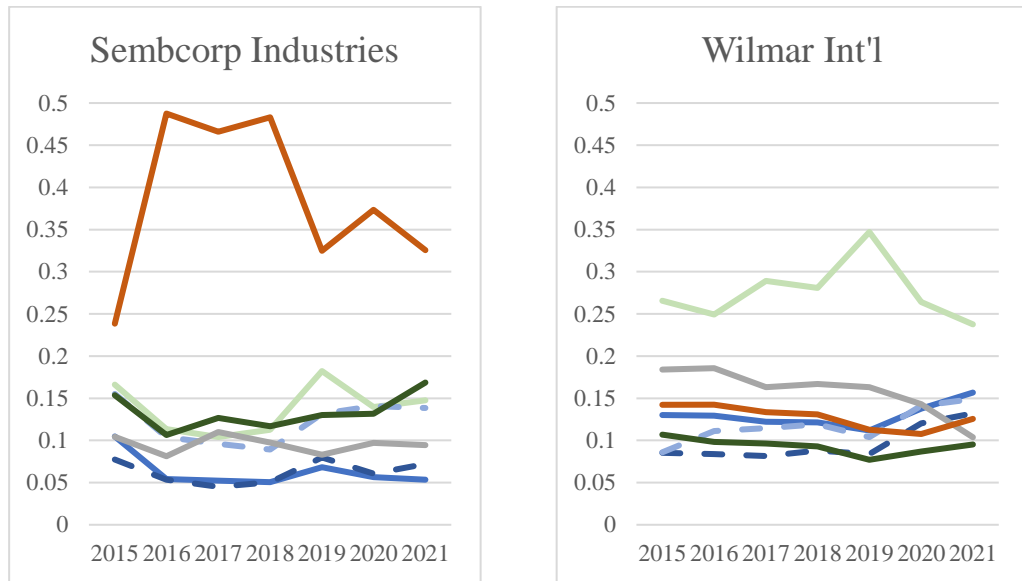


Figure 22: Distribution of topics by year for selected companies

The distribution of topics can be understood through the signalling and legitimacy theories.

The message the companies are trying to communicate (signal) is based on the industry they operate in. For example, Keppel and Sembcorp have operations in Offshore and Marine. A corruption scandal badly plagued this industry in Singapore's context in the early 2010s – SembMarine and Keppel Offshore and Marine were found to have made bribes. Therefore, the relative emphasis on disclosing governance topics could be related to the companies trying to clear their bad reputation.

Companies also seek to gain society's approval. For example, Comfortdelgro, one of Singapore's largest transportation groups, was focused on customer-centric disclosures before pivoting sharply to climate-themed disclosures for their 2021 Sustainability Report. This reflects their communication strategy – to be customer-centric. Wilmar is an agribusiness group. Their emphasis on resources and supply chains is an attempt at gaining legitimacy.

Readability of Sustainability Reports

The metric used is the Flesch Reading Ease Score. The results are presented across two dimensions – the company and the years. The top three (3) and bottom three (3) companies in terms of readability are featured in Table 19. The readability scores are presented in Table 17. The full list of readability scores is enclosed in **Appendix IV**.

Year	Count of Reports	Average RE	Min. RE	Max. RE	Std. Dev. RE
Overall	200	9.107	-15.434	37.747	8.803

Table 17: Summary Statistics of Readability Scores for Sustainability Reports

The readability score of sustainability reports is low. This was observed by Courtis (1995), Courtis (1998), Stanton and Stanton (2002), and Li (2008a). The average person will struggle to decode sustainability reports (Farewell et al., 2014) because of the complexity of the terms used. This is evidenced even by using a measure which examines the average sentence length and average syllable per word.

With more requirements being introduced and more disclosure content, the readability of reports is deteriorating. This is evidenced in Table 18.

Year	Count of Reports	Average RE	Min. RE	Max. RE	Std. Dev. RE
2015	27	14.111	-2.996	28.710	8.451
2016	27	11.535	-6.993	22.733	7.072
2017	29	10.743	-4.408	28.616	7.193
2018	29	9.144	-15.434	37.747	11.154
2019	29	7.129	-11.893	25.938	7.757
2020	29	6.667	-8.954	22.529	7.779
2021	30	5.073	-7.450	32.657	8.840

Table 18: Summary Statistics of Readability Scores for Sustainability Reports by Financial Year

Top 3 in Readability	RE	Pages	Words
DAIRY FARM INT'L HOLDINGS LTD (2018)	37.747	2	790
DAIRY FARM INT'L HOLDINGS LTD (2021)	32.657	2	1,185
YANGZIJIANG SHIPBLDG HLDGS LTD (2015)	28.710	4	1,827
Bottom 3 in Readability	RE	Pages	Words
OVERSEA-CHINESE BANKING CORP (2018)	-15.434	18	13,156
HONGKONG LAND HOLDINGS LIMITED (2019)	-11.893	63	21,144
HONGKONG LAND HOLDINGS LIMITED (2020)	-8.954	84	28,688

Table 19: Top and Bottom 3 Reports in terms of readability

Evidence suggests that the longer a sustainability report is in terms of the number of words, the less readable it is. Figure 23 below demonstrates that there is a statistically significant relationship between (p-value = 0.015) the number of words and the Flesch Readability Ease Score, albeit the relationship is not linear.

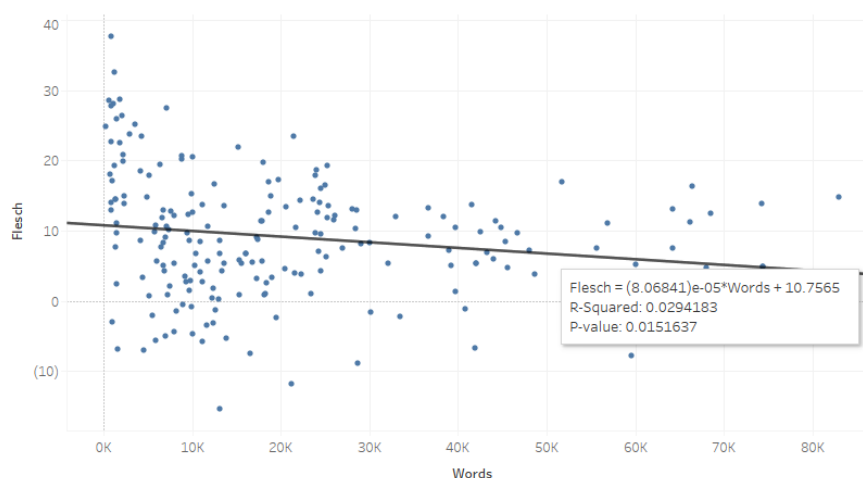


Figure 23: Readability Scores by Number of Words in Sustainability Reports

Sentiments of Sustainability Reports

As discussed in the section on the derivation of dependent variables, the sentiment score is different for every lexicon. This is because each lexicon has a different composition of words, resulting in different words picked out in the document, leading

to a variance in scores. Using Keppel Corporation Limited's 2015 Sustainability Report, this variance can be visualised in Figure 24 and Figure 25.

The sentiment scores by page generally appear positive. Only five out of the 96 pages were recorded as having an average negative sentiment score. A similar trend is observed in other sustainability reports as well. For a page to have a positive sentiment score, it shall have more positive words than negative words. Therefore, the sentiment scores being overall positive supports Harymawan et al. (2020)'s observation that companies appear to use a choice of words that contain more positive sentiments.

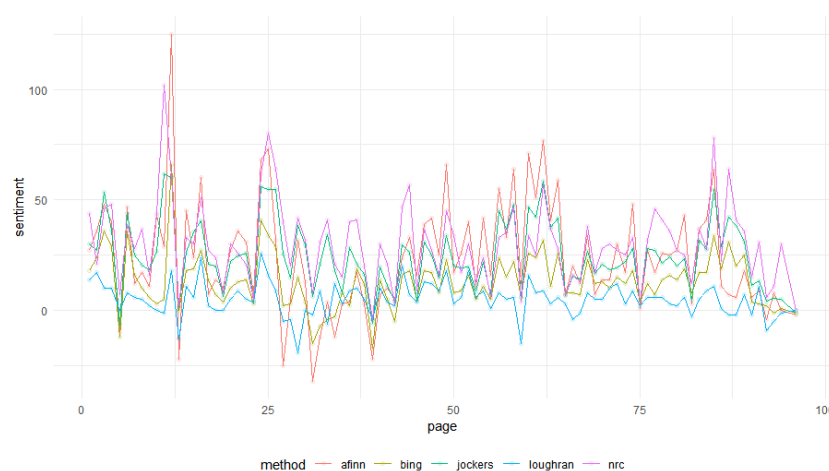


Figure 24: Sentiment Score, by Page, in Keppel's 2015 Sustainability Report

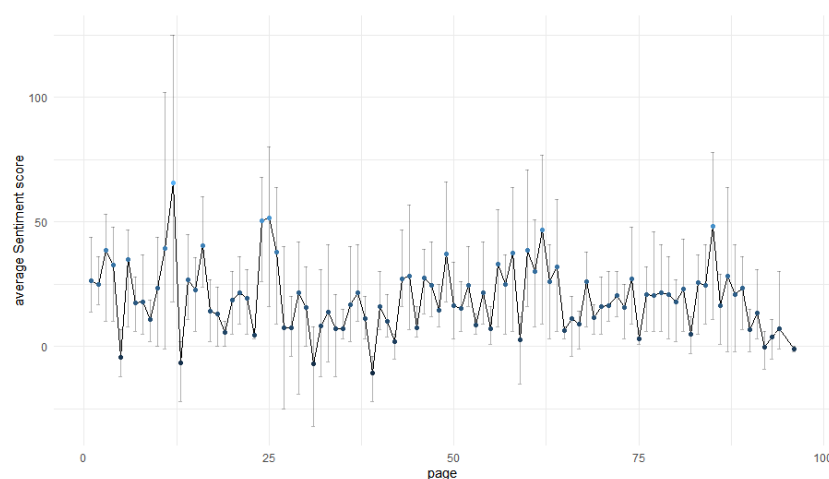


Figure 25: Average Sentiment Score, with Error Bars representing maximum and minimal values, by Page in Keppel's 2015 Sustainability Report

It can also be noted that sustainability reports tend to start with a more positive language before tapering down towards the end. The distribution of words with positive and negative sentiments is generally consistent across the majority of the 200 sentiment reports in the sample.

The average sentiment per page has been consistent throughout the years, although it has decreased slightly but returned to 2015 levels for the financial years ending in 2021. This concurs with Boiral (2013)'s finding that more positive language is used in years where there is bad news – e.g., during the COVID-19 pandemic.

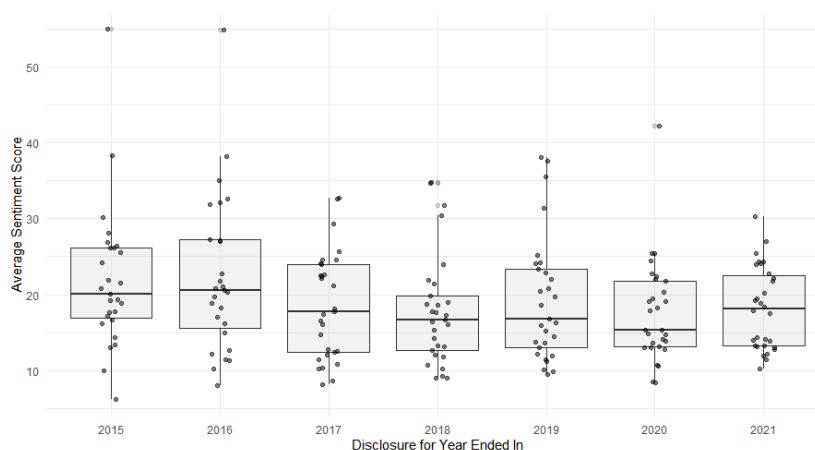


Figure 26: Average Sentiment Score, adjusted for number of pages, by Year

The summary statistics for average sentiment score, adjusted for the number of pages in each sustainability report, is presented in Table 20 overall, Table 21 by year, and Table 22 for the top and bottom three companies. The standard deviation of sentiment scores within companies is noticeably lower, and despite the occasional exceptions, most companies have a standard deviation of average sentiment scores between 1 to 4.

Year	Count of Reports	Average S	Min. S	Max. S	Std. Dev. S
Overall	200	19.344	6.219	54.960	7.981

Table 20: Summary Statistics of Sentiment Scores for Sustainability Reports

Year	Count of Reports	Average S	Min. S	Max. S	Std. Dev. S
2015	27	21.897	6.219	54.960	9.426
2016	27	22.387	8.007	54.900	10.368
2017	29	18.566	8.123	32.754	7.055
2018	29	17.853	8.949	34.743	7.324
2019	29	19.281	9.495	38.130	8.218
2020	29	17.679	8.381	42.213	6.800
2021	30	18.172	10.171	30.339	5.441

Table 21: Summary Statistics of Sentiment Scores for Sustainability Reports by Financial Year

Top 3 Companies in Sentiments	S	Std. Dev. S
OVERSEA-CHINESE BANKING CORP	29.005	10.940
SEMBICORP INDUSTRIES LTD	28.347	5.232
JARDINE MATHESON HLDGS LTD	25.808	8.052
Bottom 3 Companies in Sentiments	S	Std. Dev. S
VENTURE CORPORATION LIMITED	12.324	3.308
UOL GROUP LIMITED	12.246	2.849
WILMAR INTERNATIONAL LIMITED	10.827	3.723

Table 22: Top and Bottom 3 Reports in terms of sentiment scores

Summary Statistics of Financial Information

The three financial measures which are used in this analysis are the Return on Assets (ROA), Market Value (MV), and Zmijewski Score (ZS). These are presented separately by year.

Return on Assets (ROA)

Year	Count of Reports	Average ROA	Min. ROA	Max. ROA	Std. Dev. ROA
Overall	200	4.65%	-11.36%	19.35%	4.43%

Table 23: Summary Statistics of ROA for Companies

Year	Count of Reports	Average ROA	Min. ROA	Max. ROA	Std. Dev. ROA
2015	27	5.79%	0.97%	19.35%	4.31%
2016	27	5.24%	0.85%	16.58%	3.83%
2017	29	6.10%	0.68%	17.61%	4.64%
2018	29	5.33%	0.96%	17.17%	3.75%
2019	29	4.79%	0.44%	18.34%	3.62%
2020	29	2.69%	-8.10%	17.61%	5.36%
2021	30	2.82%	-11.36%	14.74%	4.30%

Table 24: Summary Statistics of ROA for Companies by Financial Year

Market Value (MV)

Year	Count of Reports	Average MV	Min. MV	Max. MV	Std. Dev. MV
Overall	200	15,321	966	88,142	16,849

Table 25: Summary Statistics of Market Value for Companies

Year	Count of Reports	Average MV	Min. MV	Max. MV	Std. Dev. MV
2015	27	13,585	966	65,652	14,495
2016	27	15,569	1,394	66,296	15,847
2017	29	16,755	1,600	79,474	19,982
2018	29	15,713	2,040	70,932	17,498
2019	29	12,984	1,412	54,185	13,800
2020	29	15,832	3,041	76,500	17,705
2021	30	16,658	3,207	88,142	18,818

Table 26: Summary Statistics of Market Value for Companies by Financial Year

Zmijewski Score (ZS)

Year	Count of Reports	Average ZS	Min. ZS	Max. ZS	Std. Dev. ZS
Overall	200	-1.808	-4.146	0.833	1.270

Table 27: Summary Statistics of Zmijewski Score for Companies

Year	Count of Reports	Average ZS	Min. ZS	Max. ZS	Std. Dev. ZS
2015	27	-1.902	-3.648	0.768	1.236
2016	27	-1.907	-3.849	0.756	1.257
2017	29	-2.004	-4.146	0.778	1.269
2018	29	-1.858	-3.740	0.783	1.308
2019	29	-1.806	-3.916	0.785	1.326
2020	29	-1.592	-3.824	0.833	1.336
2021	30	-1.610	-3.830	0.821	1.224

Table 28: Summary Statistics of Zmijewski Score for Companies by Financial Year

Interpreting Regression Results

The analysis will be broken up into three subsections: an analysis of the Return on Assets – the company’s ability to utilise its resources, the Market Value – influenced by external perception; and the Zmijewski Score – a measure of the company as a going concern.

The following research questions will be discussed:

RQ 2. How do these topics identified in RQ 1 relate to CFP?

RQ 3. How do a sustainability report's readability and sentiment score relate to corporate financial performance?

Analysis on Return on Assets

The return on assets is a profitability ratio that indicates the amount of profit a company can generate in a financial period based on the amount invested in assets at the end of the financial period. This is a measure of the company’s internal capability.

The results in Table 29 suggest the following:

- The extent of community-related disclosures, at a 0.06% significant level, has a positive relationship with the company’s return on assets. This means that the more a company discloses what it is doing for the community, the higher the return on assets will be – despite a marginal effect with $\beta = 4.2 * 10^{-6}$.

This result supports the legitimacy theory – that companies need to act in congruence with society to uphold their business activities (O’Donovan, 2002) and that CSR disclosures influence society’s perceptions about the company and act as an empowering effect to better the firm’s financial performance (Hooghiemstra, 2002).

- There is no statistically significant relationship between the other disclosure areas and a company's return on assets. Although this is not conclusive, this may suggest that other factors do not influence a company's ability to operate efficiently to produce returns for component companies in the STI.
- A sustainability report's readability and sentiment score does not significantly influence the company's return on assets. By extension, the sustainability report, as a communication medium of the company, does not have a statistically significant relationship with the company's internal operational efficiency.

	RQ 2		RQ 3		Combined	
	β	<i>p</i> -value	β	<i>p</i> -value	β	<i>p</i> -value
<i>Intercept</i>	6.33 E ⁻²	< 2 E ⁻¹⁶ (***)	5.08 E ⁻²	3.79 E ⁻⁸ (***)	8.97 E ⁻²	2.81 E ⁻¹³ (***)
<i>S: Community</i>	4.32 E ⁻⁶	0.000695 (***)	-	-	4.93 E ⁻⁶	0.00014 (***)
<i>E: Resources</i>	-4.82 E ⁻⁷	0.842	-	-	1.74 E ⁻⁷	0.942
<i>S: Customers</i>	-2.87 E ⁻⁶	0.099 (.)	-	-	-3.02 E ⁻⁶	0.080
<i>S: Employees</i>	-5.49 E ⁻⁶	0.260	-	-	-6.38 E ⁻⁶	0.190
<i>Methodology</i>	1.71 E ⁻⁸	0.997	-	-	-4.06 E ⁻⁶	0.426
<i>G: Governance</i>	-4.51 E ⁻⁸	0.976	-	-	3.53 E ⁻⁷	0.815
<i>E: Climate</i>	-1.54 E ⁻⁶	0.339	-	-	5.42 E ⁻⁹	0.998
<i>Readability</i>	-	-	9.26 E ⁻⁵	0.796	-4.97 E ⁻⁴	0.157
<i>Sentiment</i>	-	-	-2.65 E ⁻⁴	0.503	-9.19 E ⁻⁴	0.0323 (*)
<i>F-Statistic, (p-value)</i>	5.671 (5.73 E ⁻⁶) ***		0.259 (0.772)		5.254 (2.17 E ⁻⁶) ***	
Adjusted R ²	0.141		-0.0075		0.161	

Table 29: Results from Multivariate Regression against Return on Assets

Analysis on Market Value

The market value indicates investors' perceptions about a company's business prospects. These values are readily available for the sample companies, given that they are all listed on the SGX. The operative word is 'perception', based on assumably publicly available information. This means that disclosure documents – i.e., financial and sustainability reports are expected to have an influence.

For the analysis of market value, Ohlson's model (Ohlson, 1995), as used in Loh et al. (2017), is used as a baseline model. This includes the variables BV_t , $EARN_t$, and $EARN_t * NEG_t$.

From Table 30, the following can be observed:

- The results support Ohlson's model in all scenarios – that the book value, BV_t , and exceptional earnings, $EARN_t$, has a statistically significant relationship with the market value four months after the financial period end, MV_{t+4} .

The financial indicators currently still have the most statistically significant influence on investors' valuation of the company.

- More disclosures on the methodology used was seen with a higher market value. The null hypothesis is rejected with a 0.8% statistical significance – indicating a statistically significant positive relationship between these two variables. There are two ways of interpreting this:
 - Firstly, companies with a larger market value are naturally making more extensive disclosures on the methodology used. Given their larger size, it is within expectations that they have more resources and a more considerable ability to do so.

- In the context of other disclosure themes, it is essential to Singapore investors that the regulator's requirements are met. This is evident in that the coefficient for the extent of methodology-related disclosures ($\beta = 2.79$) is the highest among all other topics.

Understanding this through the legitimacy theory, companies that are more compliant are more positively perceived by investors.

- The other two themes of disclosures that investors responded to were Community (positively) and Resources (negatively).

As observed in the analysis of ROA, community-related disclosures were the only theme of disclosures to have an impact on the ROA. This understanding can be supplemented with the stakeholder theory – in that investors believe that a company which serves its stakeholders – particularly the community it operates in, is poised to perform well, and is thus valued more. In Singapore's context, it may validate the “all of society” approach championed by the government in encouraging firms to contribute to social advancement.

For resources, the negative relationship could reveal that more disclosure reveals that current initiatives are insufficient or make investors feel less confident in the company's ability to perform. Moreover, it could also reflect the Friedman doctrine embedded in investors' biases. For instances, the more a company does regarding conserving resources results in additional costs, reducing the company's perceived profitability in the short term, and negatively affecting investors' confidence. These results alone are inclusive.

- In an individual model with readability and sentiment scores – the sentiment score has a statistically significant negative relationship with the market value. The more positive the words used were, the lower the observed market value. This tends towards the view Hrasky (2012) and Boiral (2013) held – more positive words were used when negative events occurred.

	RQ 2		RQ 3		Combined	
	β	p-value	β	p-value	β	p-value
<i>Intercept</i>	1.06 E ³	0.371	6629.45	0.000627 (***)	2.38 E ³	0.333
<i>BV_t</i>	3.219 E ⁻¹	0.000288 (***)	0.342	0.000228 (***)	3.384 E ⁻¹	0.000185 (***)
<i>EARN_t</i>	6.037	6.92 E ¹¹ (***)	6.47	1.41 E ⁻¹¹ (***)	5.974	1.72 E ⁻¹⁰ (***)
<i>EARN_t * NEG_t</i>	-5.427	0.0201 (*)	-6.31	0.00862 (**)	-5.433	0.0203 (*)
<i>S: Community</i>	5.63 E ⁻¹	0.0418 (*)	-	-	5.60 E ⁻¹	0.0464 (*)
<i>E: Resources</i>	-1.66	0.00144 (**)	-	-	-1.65	0.00167 (**)
<i>S: Customers</i>	3.29 E ⁻¹	0.393	-	-	2.89 E ⁻¹	0.0455
<i>S: Employees</i>	-7.72 E ⁻²	0.940	-	-	-2.43 E ⁻¹	0.817
<i>Methodology</i>	2.79	0.00883 (**)	-	-	2.63	0.0188 (*)
<i>G: Governance</i>	-3.34 E ⁻¹	0.293	-	-	-32.31 E ⁻¹	0.486
<i>E: Climate</i>	-1.73 E ⁻¹	0.627	-	-	-4.51 E ⁻²	0.906
<i>Readability</i>	-	-	4.702	0.951	48.37	0.533
<i>Sentiment</i>	-	-	-183.58	0.0312 (*)	-89.06	0.346
<i>F-Statistic, (p-value)</i>	51.12 (< 2.2 E ⁻¹⁶) ***		79.46 (< 2.2 E ⁻¹⁶) ***		43.19 (< 2.2 E ⁻¹⁶) ***	
Adjusted R ²	0.735		0.703		0.734	

Table 30: Results from Multivariate Regression against Market Value

Analysis on Zmijewski Score

Unlike the company's ROA, the Zmijewski score considers the company's profitability, solvency, and liquidity. Based on the weighting of the Zmijewski score, the profitability and solvency ratios are the most decisive factors for the Zmijewski score. The score (Zmijewski, 1984) proposed the score as a bankruptcy model. It can also be interpreted as the company's financial viability – and the ability to exist as a going concern. Higher scores are indicative of a lower probability of default.

In Table 31, the following is observed:

- The extent of climate and governance disclosures had more positive disclosures have a statistically significant positive relationship with the Zmijewski score. i.e., the more a firm discloses climate-related activities, the more financially viable it is. This is encouraging empirical support for the modern view of the triple bottom line of profits, people, and the planet as a barometer for a company's success.

This finding is supported by the resource theory and transaction cost economics. In recent years, companies which do more for the climate, build more resources, present themselves as more responsible, and internally take the posture of being more financially viable in the long run.

- The readability aspect of a sustainability report has the most statistically significant relationship to the Zmijewski score compared to other financial metrics. While it is not statistically significant enough for us to reject the null hypothesis, this hints that the readability of such reports and the ability to be understood through these reports is reflected in the company's long-term financial viability and view of itself. The signalling theory supports this inference.

	RQ 2		RQ 3		Combined	
	β	p-value	β	p-value	β	p-value
<i>Intercept</i>	-1.949	$< 2 \text{ E}^{-16}$ (***)	-2.676	$< 2 \text{ E}^{-16}$ (***)	-2.748	2.79 E^{-15} (***)
<i>S: Community</i>	4.55 E^{-5}	0.201	-	-	1.86 E^{-5}	0.601
<i>E: Resources</i>	5.15 E^{-5}	0.453	-	-	2.68 E^{-5}	0.690
<i>S: Customers</i>	5.75 E^{-5}	0.243	-	-	5.97 E^{-5}	0.214
<i>S: Employees</i>	-2.07 E^{-4}	0.135	-	-	-2.01 E^{-4}	0.140
<i>Methodology</i>	-3.09 E^{-4}	0.0273 (*)	-	-	-1.72 E^{-4}	0.230
<i>G: Governance</i>	1.25 E^{-4}	0.00334 (**)	-	-	1.26 E^{-4}	0.00333 (**)
<i>E: Climate</i>	2.18 E^{-4}	3.55 E^{-6} (***)	-	-	1.81 E^{-4}	0.000251 (***)
<i>Readability</i>	-	-	0.162	0.103	2.91 E^{-2}	0.00330 (**)
<i>Sentiment</i>	-	-	0.0372	0.000818 (***)	2.09 E^{-2}	0.0815 (.)
<i>F-Statistic, (p-value)</i>	6.474 (7.40 E^{-7}) ***		7.111 (0.001043) **		6.612 (3.34 E^{-8}) ***	
Adjusted R^2	0.162		0.0578		0.2024	

Table 31: Results from Multivariate Regression against Zmijewski Score

RQ 4. Are there material differences in the relationship between CSR disclosures and CFP at this point, vis-à-vis prior studies conducted by Tsang, 1998 and Loh et al., 2017?

In the context of Singapore-specific studies, there were three:

1. Tsang (1998) analysed 17 companies' disclosures from 1986 to 1995. Tsang (1998) found that disclosures were primarily qualitative, and measured disclosure as a percentage of sentences in sustainability reports. 'Community involvement' disclosures were significantly higher than environmental disclosures throughout the observational period. In 1995, 0.24% of sentences in the reports were related to the environment, whilst 5.37% of sentences were related to community involvement.

Regarding the proportions, environmental-related disclosures were significantly higher relative to community-related disclosures. For example, disclosures solely related to community-related initiatives (Topic 1: Community) are half of the environmental-related disclosures (Topic 2: Resources, Topic 7: Climate). This is motivated by regulatory requirements from SGX but also possibly explained through the institutional and stakeholder theory – where companies are more pressed to report on environmental issues important to stakeholders and society.

2. Loh et al. (2017) used Ohlson's model as a baseline and investigated the statistical significance of a sustainability reporting score on the market value. In this study, Loh et al. (2017) used the mere presence of information relating to an area of sustainability reporting to score sustainability reports. The study found, with statistical significance, that sustainability disclosure is positively related to the market value of a firm and that the better the quality of sustainability reporting, the stronger the linkage.

In this study, it is shown that there is a statistically significant relationship between the extent of disclosures of specific topics and the market value. The effects are constrained to the methodology, community (positive relationship), and resources (negative relationship).

Both studies confirm that sustainability disclosures have a statistically significant relationship with the firm's market value.

Conclusion

Prior CSR studies either used platform-generated scores or rudimentary inputs for assessing a company's CSR disclosures. Where NLP techniques were used in recent studies, these were either used in isolation or on companies listed in markets other than Singapore. This study proposes a robust methodology for quantifying and assessing a company's CSR disclosures using NLP techniques applied directly to sustainability reports for companies listed on the SGX.

The results have allowed us to form inferences about statistically significant or insignificant relationships between CSR disclosures – topics, readability, sentiments – and corporate financial performance for companies which are a part of the STI.

This study found that specific topics of sustainability disclosures had statistically significant relationships with different financial performance factors. For example, community-related disclosures had a positive relationship with the ROA and MV. Whilst methodology-related disclosures had a positive relationship with MV, resource-related disclosures had a negative relationship with MV. Lastly, governance and climate-related disclosures had a positive relationship with the Zmijewski score, indicating that these factors contributed to a firm's financial viability.

Despite this, there was no conclusive evidence that readability nor the sentiments of words used in sustainability reports impacted financial performance.

Perspective is critical. If GDP and profitability remain the prime metrics (and products are evaluated by price and feature comparison without proper life-cycle analysis and environmental “weighting”, then sustainability will always be a secondary consideration (Rodgers et al., 2013).

The findings of this study, in conjunction with encouraging empirical evidence from other studies, suggest that there are benefits of CSR disclosures and there are positive effects on financial performance. It is hoped that this study will contribute suggestions on how CSR disclosures can be assessed and enable an understanding of what CSR factors are important for stakeholders to encourage companies to adopt more of such positive initiatives.

Limitations and Opportunity for Future Studies

Similar to Rodgers et al. (2013), there was a limit on the number of companies analysed. Because of the smaller sample size, performing industry differentiation was impossible. For example, it may be worth focusing on the disclosure effects for critical industries identified by SGX has a high impact – (i) financial, (ii) agriculture, (iii) energy, (iv) materials and buildings, and (v) transportation. These are the five industries stipulated by SGX which are required to present mandatory climate-related disclosures in line with the TCFD’s recommendations. Similarly, it may also be possible to examine differences between Government-Linked Companies, Family Businesses, and others, as Loh et al. (2017) did. A study for listed companies in each of these industries can be conducted with a more extensive sample size – focusing on all companies listed, rather than those included as part of the STI.

In this study, all reports across the observation period of 2015 to 2021 were treated as part of the same sample space. SGX launched several sustainability regulations in this period which may result in different conclusions being drawn. For example, implementing the TCFD’s recommendations for climate-related disclosures is a significant event. In subsequent periods, when companies begin to make such

disclosures, examining the effects on the relationships identified here pre- and post-implementation of the requirements are meaningful. The results in this study provide a preliminary basis.

The analysis relating to market value raises questions about Singapore investors' perceptions. There were inferences about investors' perception of the importance of methodology disclosures, community-related initiatives, and resource-related topics and their effect on the market value. These should be validated or disproved through a survey of investors. The mediating and moderating factors influencing the market value can be more detailed.

In the same vein, while this study provides a basis for observing the relationships, it does not conclude the causal effects. Instead, this study merely identifies statistically significant relationships between CSR disclosure variables and CFP without any suggestions on causality. To help business leaders understand the causal effects of CSR disclosures in present times, an update to the conceptual framework – similar to that of Richardson et al. (1999) – can be proposed. This will help stakeholders understand if there are new influences based on how sustainability has become important to society amid a climate emergency.

References

- Abu Bakar, A. S., & Ameer, R. (2011). Readability of corporate social responsibility communication in Malaysia. *Corporate Social Responsibility and Environmental Management*, 18(1), 50-60.
- Arco-Castro, L., López-Pérez, M. V., Pérez-López, M. C., & Rodríguez-Ariza, L. (2020). How market value relates to corporate philanthropy and its assurance. The moderating effect of the business sector. *Business Ethics: A European Review*, 29(2), 266-281.
- Aggarwal, P. (2013a). Impact of sustainability performance of company on its financial performance: A study of listed Indian companies. *Global Journal of Management and Business Research (C: Finance) Volume*, 13.
- Aggarwal, P. (2013b). Sustainability reporting and its impact on corporate financial performance: A literature review. *Indian Journal of Commerce and Management Studies*, 4(3), 51
- Aguilera - Caracuel, J., & Guerrero - Villegas, J. (2018). How corporate social responsibility helps MNEs to improve their reputation. The moderating effects of geographical diversification and operating in developing regions. *Corporate social responsibility and environmental management*, 25(4), 355-372.
- Aguilera, R. V., & Jackson, G. (2003). The cross-national diversity of corporate governance: Dimensions and determinants. *Academy of Management Review*, 28, 447-465.
- Aravindan, A. (2021, December 15). Singapore Exchange to mandates climate, board diversity disclosures. Reuters. <https://www.reuters.com/business/sustainable-business/singapore-exchangemandates-climate-board-diversity-disclosures-2021-12-15/>
- Artiach, T., Lee, D., Nelson, D., & Walker, J. (2010). The determinants of corporate sustainability performance. *Accounting & Finance*, 50(1), 31-51.

Back, B., Toivonen, J., Vanharanta, H., & Visa, A. (2001). Comparing numerical data and text information from annual reports using self-organizing maps. *International journal of accounting information systems*, 2(4), 249-269.

Backhaus, K. B., Stone, B. A., & Heiner, K. (2002). Exploring the relationship between corporate social performance and employer attractiveness. *Business & Society*, 41(3), 292-318.

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

Balakrishnan, R., Qiu, X. Y., & Srinivasan, P. (2010). On the predictive ability of narrative disclosures in annual reports. *European Journal of Operational Research*, 202(3), 789-801.

Basel Committee on Banking Supervision. (2015). Guidelines corporate governance principles for banks. Bank for International Settlements, 1.

Berman, S. L., Wicks, A. C., Kotha, S., & Jones, T. M. (1999). Does stakeholder orientation matter? The relationship between stakeholder management models and firm financial performance. *Academy of Management journal*, 42(5), 488-506.

Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python: analyzing text with the natural language toolkit. " O'Reilly Media, Inc."

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77-84.

Bogert, J. (1985). In defense of the Fog Index. *Business Communication Quarterly*, 48(2), 9-12.

doi:10.1177/108056998504800203

Boiral, O. (2013). Sustainability reports as simulacra? A counter-account of A and A+ GRI

reports. *Accounting, Auditing & Accountability Journal*, 26, 1036-1071. doi:10.1108/AAAJ-04-2012-00998

Bowen, H. R. (1953). *Social responsibilities of the businessman*. New York, NY: Harper.

Bowen, K. J., Craddock-Henry, N. A., Koch, F., Patterson, J., Häyhä, T., Vogt, J., & Barbi, F. (2017). Implementing the “Sustainable Development Goals”: towards addressing three key governance

Brammer, S., Millington, A., & Rayton, B. (2007). The contribution of corporate social responsibility to organizational commitment. *International Journal of Human Resource Management*, 18(10), 1701–1719.

Brogi, M., & Lagasio, V. (2019). Environmental, social, and governance and company profitability: Are financial intermediaries different?. *Corporate Social Responsibility and Environmental Management*, 26(3), 576-587.

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.

Butler, M., & Kešelj, V. (2009, May). Financial forecasting using character n-gram analysis and readability scores of annual reports. In *Canadian Conference on Artificial Intelligence* (pp. 39-51). Springer, Berlin, Heidelberg.

Cao, J., Xia, T., Li, J., Zhang, Y., & Tang, S. (2009). A density-based method for adaptive LDA model selection. *Neurocomputing*, 72(7-9), 1775-1781.

C.J. van Rijsbergen, S.E. Robertson and M.F. Porter, 1980. *New models in probabilistic information retrieval*. London: British Library. (British Library Research and Development Report, no. 5587).

Carroll, A. B., & Shabana, K. M. (2010). The business case for corporate social responsibility: A review of concepts, research and practice. *International journal of management reviews*, 12(1), 85-105.

Campbell, J. L., Chen, H., Dhaliwal, D. S., Lu, H. M., & Steele, L. B. (2014). The information content of mandatory risk factor disclosures in corporate filings. *Review of Accounting Studies*, 19(1), 396-455.

Chang, Y. K., Oh, W. Y., Park, J. H., & Jang, M. G. (2017). Exploring the relationship between board characteristics and CSR: Empirical evidence from Korea. *Journal of Business Ethics*, 140(2), 225-242.

Chia K. H., Wee, E. (2021, December 28). Enhanced disclosures on climate-related information & board diversity policy for SGX-listed companies. Lexology. Retrieved from <https://www.lexology.com/library/detail.aspx?g=1b4ea3d2-1330-4f35-8e5f-447f181e3e64> on 2022, April 13

Chin, W. W. 1998. Issues and opinion on structural equation modeling. *MIS Quart.* 22 7-16.

Chin, W. W., B. L. Marcolin, P. R. Newsted, 2003. A partial least squares latent variable modelling approach for measuring interaction effects: Results from a Monte Carlo simulation study and an electronic mail emotion/adoption study. *Inform. Syst. Res.* 14: 189-217.

Ching, H. Y., Gerab, F., & Toste, T. H. (2017). The quality of sustainability reports and corporate financial performance: Evidence from Brazilian listed companies. *Sage Open*, 7(2), 2158244017712027.

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.

Chowdhury, R. H., Choi, S., Ennis, S., & Chung, D. (2019). Which dimension of corporate social responsibility is a value driver in the oil and gas industry?. *Canadian Journal of Administrative Sciences/Revue Canadienne des Sciences de l'Administration*, 36(2), 260-272.

Clarkson, P. M., Li, Y., Richardson, G. D., & Vasvari, F. P. (2008). Revisiting the relation between environmental performance and environmental disclosure: An empirical analysis. *Accounting, organizations and society*, 33(4-5), 303-327.

Clatworthy, M., & Jones, M. J. (2001). The effect of thematic structure on the variability of annual report readability. *Accounting, Auditing & Accountability Journal*.

Clatworthy, M., & Jones, M. J. (2003). Financial reporting of good news and bad news: evidence from accounting narratives. *Accounting and business research*, 33(3), 171-185.

Clatworthy, M. A., & Jones, M. J. (2006). Differential patterns of textual characteristics and company performance in the chairman's statement. *Accounting, Auditing & Accountability Journal*.

Conneau, A., Kiela, D., Schwenk, H., Barrault, L., & Bordes, A. (2017). Supervised learning of universal sentence representations from natural language inference data. *arXiv preprint arXiv:1705.02364*.

Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2011). Signaling theory: A review and assessment. *Journal of management*, 37(1), 39-67.

Cormier, D., Magnan, M., & Van Velthoven, B. (2005). Environmental disclosure quality in large German companies: economic incentives, public pressures or institutional conditions?. *European accounting review*, 14(1), 3-39.

Cornelius, N., Wallace, J., & Tassabehji, R. (2007). An analysis of corporate social responsibility, corporate identity and ethics teaching in business schools. *Journal of Business Ethics*, 76(1), 117-135.

Cöster, M., Dahlin, G., & Isaksson, R. (2020). Are they reporting the right thing and are they doing it right?—A measurement maturity grid for evaluation of sustainability reports. *Sustainability*, 12(24), 10393.

Courtis, J. K. (1995). Readability of annual reports: Western versus Asian evidence. *Accounting, Auditing & Accountability Journal*.

Courtis, J. K. (1998). Annual report readability variability: tests of the obfuscation hypothesis. *Accounting, Auditing & Accountability Journal*.

Crain, S. P., Zhou, K., Yang, S. H., & Zha, H. (2012). Dimensionality reduction and topic modeling: From latent semantic indexing to latent dirichlet allocation and beyond. In *Mining text data* (pp. 129-161). Springer, Boston, MA.

Cupertino, S., Consolandi, C., & Vercelli, A. (2019). Corporate social performance, financialization, and real investment in US manufacturing firms. *Sustainability*, 11(7), 1836.

Dahlsrud, A. (2008): "How Corporate Social Responsibility Is Defined: An Analysis of 37 Definitions." *Corporate Social Responsibility and Environmental Management.*, John Wiley & Sons, 15 (1), 1–13.

Dangelico, R. M., & Pontrandolfo, P. (2010). From green product definitions and classifications to green option matrix. *Journal of Cleaner Production*, 18(16/17), 1608–1628.

De Franco, G., Hope, O. K., Vyas, D., & Zhou, Y. (2015). Analyst report readability. *Contemporary Accounting Research*, 32(1), 76-104.

Delmas, M., & Toffel, M. W. (2004). Stakeholders and environmental management practices: An institutional framework. *Business Strategy and the Environment*, 13, 2009–2222.

Deveaud, R., SanJuan, E., & Bellot, P. (2014). Accurate and effective latent concept modeling for ad hoc information retrieval. *Document numérique*, 17(1), 61-84.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

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

Elkington, J., & Rowlands, I. H. (1999). Cannibals with forks: The triple bottom line of 21st century business. *Alternatives Journal*, 25(4), 42.

Engelberg, J. (2008). Costly information processing: Evidence from earnings announcements. In AFA 2009 San Francisco meetings paper.

Fang, E.-L., Loh, L., Lim, L. H., Singh, A. A., Nguyen, T. T. A., Tan, S., & Thoi, V. (2022, May). Sustainability counts: Understanding sustainability reporting requirements across Asia Pacific and insights on the journey to date. PricewaterhouseCoopers and Centre for Governance and Sustainability, NUS Business School. <https://bschool.nus.edu.sg/cgs/wp-content/uploads/sites/7/2022/05/CGS-PwC-Sustainability-Counts-Asia-Pacific-Report-2022.pdf>

Farewell, S., Fisher, I., & Daily, C. (2014). The lexical footprint of sustainability reports: A pilot study of readability. In American Accounting Association Annual Meeting and Conference on Teaching and Learning in Accounting, Sarasota, FL.

Fasan, M., Marcon, C., & Mio, C. (2016). Institutional determinants of IR disclosure quality. In C. Mio (Ed.). London, UK: Integrated reporting. A new accounting disclosure, Palgrave Macmillan.

Fatoki, O. (2019). Green entrepreneurial orientation and firm performance in South Africa. *Entrepreneurship and Sustainability Issues*, 7(1), 247.

Feinerer I, Hornik K, Meyer D (2008). "Text Mining Infrastructure in R." *Journal of Statistical Software*, 25(5), 1–54. <https://www.jstatsoft.org/v25/i05/>.

Fisher, I. E., Garnsey, M. R., & Hughes, M. E. (2016). Natural language processing in accounting, auditing and finance: A synthesis of the literature with a roadmap for future research. *Intelligent Systems in Accounting, Finance and Management*, 23(3), 157-214.

Flammer, C., & Luo, J. (2017). Corporate social responsibility as an employee governance tool: Evidence from a quasi-experiment. *Strategic Management Journal*, 38(2), 163-183.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32, 221-233. doi:10.1037/h0057532

Freundlieb, M., & Teuteberg, F. (2013). Corporate social responsibility reporting-a transnational analysis of online corporate social responsibility reports by market-listed companies: contents and their evolution. *International Journal of Innovation and Sustainable Development*, 7(1), 1-26.

Friedman, M. (1970), "The social responsibility of business is to increase its profits," *New York Times Magazine*, 32(13), 122-126

FSB TFCFD (2017). Final report: recommendations of the task force on climate-related financial disclosures. Financial Stability Board Task Force on Climate-Related Disclosures. Retrieved from <https://www.fsb-tcfd.org/recommendations/> on 2022, Jul 24.

Galani, D., Gravas, E., & Stavropoulos, A. (2012). Company characteristics and environmental policy. *Business Strategy and the Environment*, 21(4), 236-247.

García-Sánchez, I. M., Martínez-Ferrero, J., & García-Meca, E. (2018). Board of directors and CSR in banking: the moderating role of bank regulation and investor protection strength. *Australian Accounting Review*, 28(3), 428-445.

Goel, S., Gangolly, J., Faerman, S. R., & Uzuner, O. (2010). Can linguistic predictors detect fraudulent financial filings?. *Journal of Emerging Technologies in Accounting*, 7(1), 25-46.

Gomez-Carrasco, P., Guillamon-Saorin, E., & Osma, B. G. (2016). The illusion of CSR: Drawing the line between core and supplementary CSR. *Sustainability Accounting, Management and Policy Journal*.

Global Reporting Initiative. (2013). An introduction to G4. The next generation of sustainability reporting. Retrieved on 2022, Jul 27 from <https://www.globalreporting.org/resource/library/GRIAn-introduction-to-G4.pdf>

Greening, D. W., & Turban, D. B. (2000). Corporate social performance as a competitive advantage in attracting a quality workforce. *Business & society*, 39(3), 254-280.

GRI, NUS, . (2022, July). Climate Reporting in ASEAN, State of Corporate Practices.

Gunning, R. (1952). *The technique of clear writing*. New York, NY: McGraw-Hill

Halliday, M. A. K. (1989). *Spoken and written language*. Language education. Oxford, England: Oxford University Press.

Harding, A. (2022, Aug 4). Public Consultation on the Draft European Sustainability Reporting Standards. AICPA and CIMA's Response. Retrieved from https://www.cimaglobal.com/Documents/Advocacy/AICPACIMA%20Sus%20Standards%20Consultation%20Response%2004%20August%202022%20FV.pdf?utm_source=mn:updt&utm_medium=email&utm_campaign=12Aug2022&utm_medium=email&utm_source=SFMC&utm_campaign=A22_Aug12&utm_content=457038

Harrison, S., & Bakker, P. (1998). Two new readability predictors for the professional writer: Pilot trials. *Journal of Research in Reading*, 21, 121-138. doi:10.1111/1467-9817.00049

Harymawan, I., Nasih, M., Ratri, M. C., Soeprajitno, R. R. W. N., & Shafie, R. (2020). Sentiment analysis trend on sustainability reporting in Indonesia: Evidence from construction industry. *Journal of Security and Sustainability Issues*, 9(3).

Henry, E. (2008). Are investors influenced by how earnings press releases are written?. *The Journal of Business Communication* (1973), 45(4), 363-407.

Hooghiemstra, R. (2000). Corporate communication and impression management—new perspectives why companies engage in corporate social reporting. *Journal of business ethics*, 27(1), 55-68.

Hrasky, S. (2012). Visual disclosure strategies adopted by more and less sustainability-driven companies. *Accounting Forum*, 36, 154-165. doi:10.1016/j.accfor.2012.02.001

Hu, M., & Liu, B. (2004, July). Mining opinion features in customer reviews. In *AAAI* (Vol. 4, No. 4, pp. 755-760).

Huang, C. -L., & Kung, F. -H. (2010). Drivers of environmental disclosure and stakeholder expectation: Evidence from Taiwan. *Journal of Business Ethics*, 96, 435–451.

Huang, C. (2022). Investors can check Singapore companies' climate disclosures with new ESG portal. The Straits Times. Retrieved 25 July 2022, from <https://www.straitstimes.com/business/companies-markets/investors-can-check-singapore-companies-climate-disclosures-with-new-esg-portal>.

Hutto, C., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of the international AAAI conference on web and social media (Vol. 8, No. 1, pp. 216-225).

Ingram, R. W., & Frazier, K. B. (1980). Environmental performance and corporate disclosure. *Journal of accounting research*, 614-622.

Jamali, D. (2008). A stakeholder approach to corporate social responsibility: A fresh perspective into theory and practice. *Journal of business ethics*, 82(1), 213-231.

Jenkins, R. (2009). 10 Corporate social responsibility. *Handbook of economics and ethics*, 69-77

Jensen, J. C., & Berg, N. (2012). Determinants of traditional sustainability reporting versus integrated reporting. An institutionalist approach. *Business Strategy and the Environment*, 21, 299–316.

Jin, N. G. (2022). Differences between Corporate Social Responsibility (CSR) and ESG (Environmental, Social, Governance) Management in the Financial Industry (Doctoral dissertation, 서울대학교 대학원).

Jockers, M. L. (2017). Syuzhet: Extract sentiment and plot arcs from Text. Retrieved from <https://github.com/mjockers/syuzhet>

Jones, T. M. (1995, July). Instrumental stakeholder theory and paradigm consensus in business and society: Advances on the methodological front. In Proceedings of the International Association for Business and Society (Vol. 6, pp. 1263-1272).

Jones, S., Frost, G., Loftus, J., & Van Der Laan, S. (2007). An empirical examination of the market returns and financial performance of entities engaged in sustainability reporting. *Australian accounting review*, 17(41), 78-87.

Kang, H., & Kim, J. (2022). Analyzing and Visualizing Text Information in Corporate Sustainability Reports Using Natural Language Processing Methods. *Applied Sciences*, 12(11), 5614.

Kasbun, N. F., Teh, B. H., & San Ong, T. (2016). Sustainability reporting and financial performance of Malaysian public listed companies. *Institutions and Economies*, 78-93.

Kaymak, T., & Bektas, E. (2017). Corporate social responsibility and governance: Information disclosure in multinational corporations. *Corporate Social Responsibility and Environmental Management*, 24(6), 555-569.

Khavesh, A., Nikhasemi, S. R., Haque, A., & Yousefi, A. (2012). Voluntary sustainability disclosure, revenue, and shareholders wealth-a perspective from Singaporean companies. *Business Management Dynamics*, 1(9), 06-12.

Kim. (2021). Current Status and direction of development of research on ESG management performance: Korea and China. *Korean-Chinese Social Science Studies*, 60, 64-83.

Kim and Park. (2021). ESG or CSR, and Corporation's Financial Performance: Literature Review and Future Research. *Journal of Strategic Management*, 24(2), 75-114.

Kincaid, J. P., Fishburne, R. P., Rogers, R. L., & Chissom, B. S. (1975, February). Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy enlisted personnel. Retrieved from <http://www.dtic.mil/dtic/tr/fulltext/u2/a006655.pdf>

Kolk, A. (2003). Trends in sustainability reporting by the Fortune Global 250. *Business strategy and the environment*, 12(5), 279-291.

Kolk, A. (2004). A decade of sustainability reporting: developments and significance. *International Journal of Environment and Sustainable Development*, 3(1), 51-64.

Korea Exchange. (2021, November). ESG definition. APA Style. Retrieved November 20, 2021, from <http://esg.krx.co.kr/contents/01/01010100/ESG01010100.jsp>

Kothari, S. P., Li, X., & Short, J. E. (2009). The effect of disclosures by management, analysts, and business press on cost of capital, return volatility, and analyst forecasts: A study using content analysis. *The Accounting Review*, 84(5), 1639-1670.

Kyaw, K., Olugbode, M., & Petracci, B. (2017). Can board gender diversity promote corporate social performance?. *Corporate Governance: The International Journal of Business in Society*.

Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2019). Albert: A lite bert for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942*.

Lawrence, A. (2013). Individual investors and financial disclosure. *Journal of Accounting and Economics*, 56(1), 130-147.

Lewis, D. D., Yang, Y., Russell-Rose, T., & Li, F. (2004). Rcv1: A new benchmark collection for text categorization research. *Journal of machine learning research*, 5(Apr), 361-397.

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

Li, F. (2010). The information content of forward-looking statements in corporate filings — A naïve Bayesian machine learning approach. *Journal of Accounting Research*, 48(5), 1049-1102.

Li, J., Pike, R., & Haniffa, R. (2008b). Intellectual capital disclosure and corporate governance structure in UK firms. *Accounting and Business Research*, 38(2), 137–159.

Liddy, E. D. (2001). Natural language processing.

Liew, T. W., Adhitya, A., & Srinivasan, R. (2014). Sustainability trends in the process industries: A text mining-based analysis. *Computers in Industry*, 65(3), 393-400.

Loh, L., Thomas, T., & Wang, Y. (2017). Sustainability reporting and firm value: Evidence from Singapore-listed companies. *Sustainability*, 9(11), 2112.

Loh, L., Thomas, T., Lee, S. P., Lim, L., Pan, H., Malek, M., Tan, S., & Nguyen, T. P. T. (2018, November). Sustainability reporting in ASEAN countries. ASEAN CSR Network and Centre for Governance and Sustainability, NUS Business School. https://www.asean-csr-network.org/c/images/Resources/Reports/2018_Sustainability_Reporting_in_ASEAN_Countries.pdf

Loh, L., & Tang, M. (2021, May). Sustainability reporting review. Singapore Exchange and Centre for Governance and Sustainability, NUS Business School. <https://bschool.nus.edu.sg/cgs/wpcontent/uploads/sites/7/2021/05/SGX-CGS-Sustainability-Reporting-Review-Report-2021.pdf>

Lohmoller, J. B. 1989. Latent Variable Path Modeling with Partial Least Squares. Springer-Verlag, New York.

Loughran, Tim, and Bill McDonald. (2011). “When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks.” *The Journal of Finance* 66 (1): 35–65. <https://doi.org/10.1111/j.1540-6261.2010.01625.x>.

Loughran, T., & McDonald, B. (2014). Measuring readability in financial disclosures. *the Journal of Finance*, 69(4), 1643-1671.

Luccioni, A., Baylor, E., & Duchene, N. (2020). Analyzing sustainability reports using natural language processing. arXiv preprint arXiv:2011.08073.

Matsui, T., Suzuki, K., Ando, K., Kitai, Y., Haga, C., Masuhara, N., & Kawakubo, S. (2022). A natural language processing model for supporting sustainable development goals: translating semantics, visualizing nexus, and connecting stakeholders. *Sustainability Science*, 17(3), 969-985.

McWilliams, A., & Siegel, D. S. (2011). Creating and capturing value: Strategic corporate social responsibility, resource-based theory, and sustainable competitive advantage. *Journal of management*, 37(5), 1480-1495.

Meng, X. H., Zeng, S. X., & Tam, C. M. (2013). From voluntarism to regulation: A study on ownership, economic performance and corporate environmental information disclosure in China. *Journal of Business Ethics*, 116, 217–232.

Merkley, K. J. (2011). More than numbers: R&D-related disclosure and firm performance (Doctoral dissertation, University of Michigan).

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

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.

Mitchell, R., Agle, B., & Wood, D. (1997). Toward a theory of stakeholder identification and salience: Defining the principle of who and what really counts. *Academy of Management Review*, 22(4), 853–886.

Modapothala, J. R., & Issac, B. (2009, January). Evaluation of corporate environmental reports using data mining approach. In 2009 International Conference on Computer Engineering and Technology (Vol. 2, pp. 543-547). IEEE.

Modapothala, J. R., Issac, B., & Jayamani, E. (2010). Appraising the corporate sustainability reports—text mining and multi-discriminatory analysis. In *Innovations in computing sciences and software engineering* (pp. 489-494). Springer, Dordrecht.

Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. *Computational intelligence*, 29(3), 436-465.

Mohammad, S. M. (2016). Sentiment analysis: Detecting valence, emotions, and other affectual states from text. In *Emotion measurement* (pp. 201-237). Woodhead Publishing.

Morgan Stanley Capital International. (2021.November). What is ESG?. APA Style. Retrieved November 20, 2021, from <https://www.msci.com/esg-101-what-is-esg>

Motwani, S. S., & Pandya, H. B. (2016). Evaluating the impact of sustainability reporting on financial performance of selected Indian companies. *International Journal of Research in IT and Management*, 6(2), 14-23.

Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. arXiv preprint arXiv:1103.2903.

O'Donovan, G. (2002). Environmental disclosures in the annual report: Extending the applicability and predictive power of legitimacy theory. *Accounting, Auditing & Accountability Journal*.

O'Leary, D. E. (2013). BIG DATA', THE 'INTERNET OF THINGS' AND THE 'INTERNET OF SIGNS. *Intelligent Systems in Accounting, Finance and Management*, 20(1), 53-65.

Oncioiu, I., Petrescu, A. G., Bîlcan, F. R., Petrescu, M., Popescu, D. M., & Anghel, E. (2020). Corporate Sustainability Reporting and Financial Performance. *Sustainability*, 12(10), 4297.

Orlitzky, M., Schmidt, F. L., & Rynes, S. L. (2003). Corporate social and financial performance: A meta-analysis. *Organization studies*, 24(3), 403-441.

Othman, I. W., Hasan, H., Tapsir, R., Rahman, N. A., Tarmuji, I., Majdi, S., ... & Omar, N. (2012, September). Text readability and fraud detection. In 2012 IEEE Symposium on Business, Engineering and Industrial Applications (pp. 296-301). IEEE.

Ott, M., Auli, M., Grangier, D., & Ranzato, M. A. (2018, July). Analyzing uncertainty in neural machine translation. In *International Conference on Machine Learning* (pp. 3956-3965). PMLR.

Panait, M., Voica, M. C., & Radulescu, I. (2014). The Activity of Capital Market Actors: Under the Sign of Social Responsibility. *Procedia Economics and Finance*, 8, 522-528.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. *arXiv preprint cs/0205070*.

Pechancová, V., Hrbáčková, L., Dvorský, J., Chromjaková, F., & Stojanović, A. (2019). Environmental management systems: an effective tool of corporate sustainability. *Entrepreneurship and Sustainability Issues*.

Pirnea, I. C., Olaru, M., & Moisa, C. (2011). Relationship between corporate social responsibility and social sustainability. *Economy Transdisciplinarity Cognition*, 14(1).

Pucheta-Martínez, M. C., & Gallego-Álvarez, I. (2018). Environmental reporting policy and corporate structures: An international analysis. *Corporate social responsibility and environmental management*, 25(5), 788-798.

Pukelis, L., Puig, N. B., Skrynik, M., & Stanciauskas, V. (2020). OSDG--Open-Source Approach to Classify Text Data by UN Sustainable Development Goals (SDGs). *arXiv preprint arXiv:2005.14569*.

Qiu, X. Y., Srinivasan, P., & Hu, Y. (2014). Supervised learning models to predict firm performance with annual reports: An empirical study. *Journal of the Association for Information Science and Technology*, 65(2), 400-413.

Reddy, K., & Gordon, L. (2010). The effect of sustainability reporting on financial performance: An empirical study using listed companies.

Rehurek, R., & Sojka, P. (2011). Gensim--python framework for vector space modelling. NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic, 3(2).

Reuters. (2021). Singapore Exchange to mandates climate, board diversity disclosures. Retrieved from <https://www.reuters.com/business/sustainable-business/singapore-exchange-mandates-climate-board-diversity-disclosures-2021-12-15/> on 2022, Sep 11.

Richardson, A. J., Welker, M., & Hutchinson, I. R. (1999). Managing capital market reactions to corporate social responsibility. *International Journal of Management Reviews*, 1(1), 17-43.

Rodgers, W., Choy, H., & Contreras, A. G. (2007). The Effects of Corporate Social Responsibility Perceptions on The Valuation of Common Stock. *Documentos de Trabajo FUNCAS*, (313), 1.

Russo, M. V., & Fouts, P. A. (1997). A resource-based perspective on corporate environmental performance and profitability. *Academy of management Journal*, 40(3), 534-559.

Rutherford, B. A. (2003). Obfuscation, textual complexity and the role of regulated narrative accounting disclosure in corporate governance. *Journal of management and governance*, 7(2), 187-210.

Salvioli, F. O. (2000). Los derechos Humanos en las convenciones internacionales de la última década del Siglo XX. *Las grandes conferencias Mundiales de la década de los 90*, 90, 11.

Schütze, H., Manning, C. D., & Raghavan, P. (2008). Introduction to information retrieval (Vol. 39, pp. 234-65). Cambridge: Cambridge University Press.

Scott, W. R. (1995). *Institutions and organizations*. Thousand Oaks, CA: Sage

Shnayder, L., Van Rijnsoever, F. J., & Hekkert, M. P. (2016). Motivations for Corporate Social Responsibility in the packaged food industry: an institutional and stakeholder management perspective. *Journal of Cleaner Production*, 122, 212-227.

Singapore Exchange. (2021, August 26). Starting with a Common Set of Core ESG Metrics. Retrieved from <https://www.sgx.com/regulation/public-consultations/20210826-consultation-paper-climate-and-diversity> on 2022, July 25

Singapore Exchange. (2022). *Sustainability reporting*. Singapore Exchange - Singapore Exchange (SGX). Retrieved from <https://www.sgx.com/regulation/sustainability-reporting> on 2022, April 13

Singhvi, S. S. (1968). Corporate disclosure through annual reports in the United States of America and India. *Journal of Finance*, 23(3), 551-552.

Smeuninx, N., De Clerck, B., & Aerts, W. (2020). Measuring the readability of sustainability reports: A corpus-based analysis through standard formulae and NLP. *International Journal of Business Communication*, 57(1), 52-85.

Smith, M., & Taffler, R. (1992a). Readability and understandability: Different measures of the textual complexity of accounting narrative. *Accounting, Auditing & Accountability Journal*, 5(4), 0-0.

Smith, M., & Taffler, R. (1992b). The chairman's statement and corporate financial performance. *Accounting & Finance*, 32(2), 75-90.

Stanton, P., & Stanton, J. (2002). Corporate annual reports: research perspectives used. *Accounting, Auditing & Accountability Journal*.

Strickland, B. (2022, July 27). AICPA & CIMA comment on ISSB drafts of sustainability disclosures. *Financial Management*. Retrieved from <https://www.fm-magazine.com/news/2022/jul/aicpa-cima-comment-issb-drafts-sustainability-disclosures.html>

Székely, N., & Vom Brocke, J. (2017). What can we learn from corporate sustainability reporting? Deriving propositions for research and practice from over 9,500 corporate sustainability reports published between 1999 and 2015 using topic modelling technique. *PloS one*, 12(4), e0174807.

Tagesson, T., Blank, V., Broberg, P., & Collin, S. O. (2009). What explains the extent and content of social and environmental disclosures on corporate websites: a study of social and environmental reporting in Swedish listed corporations. *Corporate social responsibility and environmental management*, 16(6), 352-364.

Taliento, M., Favino, C., & Netti, A. (2019). Impact of environmental, social, and governance information on economic performance: Evidence of a corporate sustainability advantage from Europe. *Sustainability*, 11(6), 1738.

TCFD (2019, June 5). Task Force on Climate-related Financial Disclosures: 2019 Status Report. disclosures. Financial Stability Board Task Force on Climate-Related Disclosures. Retrieved from <https://www.fsb.org/2019/06/task-force-on-climate-related-financial-disclosures-2019-status-report/> on 2022, Jul 23

Townsend, S., Bartels, W., & Renaut, J.-P. (2010). Reporting change. *Change*, 1-33. Retrieved from <http://www.sustainability.com/library>

Tsang, E. W. (1998). A longitudinal study of corporate social reporting in Singapore: The case of the banking, food and beverages and hotel industries. *Accounting, Auditing & Accountability Journal*.

Vogel, D. J. (2005). Is there a market for virtue?: The business case for corporate social responsibility. *California management review*. 47(4), 19-45

Waddock, S. A., & Graves, S. B. (1997). The corporate social performance–financial performance link. *Strategic management journal*, 18(4), 303-319.

Wahyuningrum, I. F. S., Oktavilia, S., Putri, N., Solikhah, B., Djajadikerta, H., & Tjahjaningsih, E. (2021). Company financial performance, company characteristics, and environmental disclosure: evidence from Singapore. In *IOP Conference Series: Earth and Environmental Science* (Vol. 623, No. 1, p. 012065). IOP Publishing.

Walker, R. (2021, May 27). Thailand leads ESG disclosure in Southeast Asia. *Fund Selector Asia*. <https://fundselectorasia.com/thailand-leads-esg-disclosure-in-southeast-asia/>

Weidenbaum, M., & Vogt, S. (1987). Takeovers and stockholders: Winners and losers. *California Management Review*, 29(4), 157-168.

Wickham, Hadley, and Romain Francois. 2016. *dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>.

Zmijewski, M. E. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting research*, 59-82.

Appendix I: Code with Narration

The code used in the analysis is presented below.

Libraries Required

The following libraries were loaded as part of the analysis.

```
library(pdftools)
library(dplyr)
library(tidytext)
library(tidyverse)
library(tm)
library(textstem)
library(quanteda.textstats)
library(topicmodels)
library(lda)
library(ldatuning)
library(ggplot2)
library(wordcloud)
library(pals)
library(SnowballC)
library(stringr)
library(quanteda)
library(reshape2)
library(lexicon)
```

Loading Sustainability Reports as Corpus

```
reports <- read.csv("sustainability_reports_sgx_sti.csv")

# Reading in Text
for (i in 1:nrow(reports)){
  if (i == 1){
    text <- pdftools::pdf_text(reports$DocumentName[i])
    # Initiate the first dataframe
    text_df <- tibble(page = 1:length(text), text = text)
    text_df$Company <- reports$Company[i]
    text_df$Year <- reports$Year[i]
  }
  else{
    text <- pdftools::pdf_text(reports$DocumentName[i])
    # Create a temporary dataframe
    text_df_temp <- tibble(page = 1:length(text), text = text)
    text_df_temp$Company <- reports$Company[i]
    text_df_temp$Year <- reports$Year[i]

    # Append to the original dataframe
    text_df <- rbind(text_df, text_df_temp)
  }
}

# Converting to a Corpus
```

```
docs <- Corpus(VectorSource(text_df$text))
```

Summarising Number of Pages

Count number of pages

```
number_pages <- text_df %>%  
  group_by(Company, Year) %>%  
  count()
```

Graphing number of pages by year (scatterplot)

```
number_pages %>%  
  mutate(  
    Year = as.character(Year)  
  ) %>%  
  ggplot(aes(x=Year, y=n)) +  
  geom_boxplot(aes(x=Year), fill = "grey", alpha = 0.2) +  
  geom_jitter(width = 0.1, alpha = 0.5) +  
  theme_minimal() +  
  ggtitle("Number of Pages by Year") +  
  ylab("Number of Pages") +  
  xlab("Disclosure for Year Ended In") +  
  theme(legend.position="none") +  
  scale_fill_brewer(palette="Blues")
```

Summarising Number of Words

Tokenisation - separating each word

```
text_tokenised <- text_df %>%  
  unnest_tokens(word, text)
```

Number of words per page

```
text_tokenised %>%  
  group_by(Company, Year, page)
```

Graphing the number of words

```
text_tokenised %>%  
  group_by(Company, Year) %>%  
  count() %>%  
  mutate(  
    Year = as.character(Year)  
  ) %>%  
  ggplot(aes(x=Year, y=n)) +  
  geom_boxplot(aes(x=Year), fill = "grey", alpha = 0.2) +  
  geom_jitter(width = 0.1, alpha = 0.5) +  
  theme_minimal() +  
  ggtitle("Number of Words by Year") +  
  ylab("Number of Words") +  
  xlab("Disclosure for Year Ended In") +  
  theme(legend.position="none")
```

Sentiments

Sentiment from lexicon - AFINN

```
afinn <- text_tokenised %>%
  inner_join(get_sentiments("afinn")) %>%
  group_by(Company, Year, page) %>%
  summarise(sentiment = sum(value)) %>%
  mutate(method = "AFINN")
```

Sentiment from lexicon - NRC

```
nrc <- text_tokenised %>%
  inner_join(get_sentiments("nrc")) %>%
  filter(sentiment %in% c("positive", "negative")) %>%
  mutate(method = "NRC") %>%
  count(Company, Year, page, sentiment) %>%
  pivot_wider(names_from = sentiment,
              values_from = n,
              values_fill = 0) %>%
  mutate(sentiment = positive - negative)
```

Sentiment from lexicon - Bing et al.

```
bing <- text_tokenised %>%
  inner_join(get_sentiments("bing")) %>%
  filter(sentiment %in% c("positive", "negative")) %>%
  mutate(method = "Bing et al.") %>%
  count(Company, Year, page, sentiment) %>%
  pivot_wider(names_from = sentiment,
              values_from = n,
              values_fill = 0) %>%
  mutate(sentiment = positive - negative)
```

Sentiment from lexicon - Loughran

```
loughran <- text_tokenised %>%
  inner_join(get_sentiments("loughran")) %>%
  filter(sentiment %in% c("positive", "negative")) %>%
  mutate(method = "Loughran") %>%
  count(Company, Year, page, sentiment) %>%
  pivot_wider(names_from = sentiment,
              values_from = n,
              values_fill = 0) %>%
  mutate(sentiment = positive - negative)
```

Sentiment from lexicon - Jockers

```
jockers <- text_tokenised %>%
  inner_join(key_sentiment_jockers) %>%
  group_by(Company, Year, page) %>%
  summarise(sentiment = sum(value)) %>%
  mutate(method = "Jockers")
```

Joining sentiments from different lexicons

```
sentiments <- afinn %>%
  mutate(afinn = sentiment) %>%
  select(-c(sentiment, method)) %>%
left_join(
  bing %>%
    mutate(bing = sentiment) %>%
    select(-c(positive, negative, sentiment))
```

```

) %>%
left_join(
  nrc %>%
    mutate(nrc = sentiment) %>%
    select(-c(positive,negative,sentiment))
) %>%
left_join(
  loughran %>%
    mutate(loughran = sentiment) %>%
    select(-c(positive,negative,sentiment))
) %>%
left_join(
  jockers %>%
    mutate(jockers = sentiment) %>%
    select(-c(sentiment,method))
)

sentiments[is.na(sentiments)] <- 0

# Calculating average sentiment across 5 lexicons
sentiments_with_average <- sentiments %>%
  mutate(
    average_sentiment = (afinn+bing+jockers+loughran+nrc)/5
  )

# Graphing Keppel's Sentiment through the Sustainability Report
sentiments %>%
pivot_longer(!c("Company","Year","page"),names_to="method",values_to="sentiment") %>%
  filter(Company == "KEPPEL CORPORATION LIMITED") %>%
  filter(Year == 2015) %>%
  ggplot(aes(x=page,y=sentiment,group=method)) +
  geom_line(aes(color=method),size=0.65) +
  geom_point(aes(color=method),alpha=0.1) +
  theme_minimal() +
  theme(legend.position="bottom") +
  scale_fill_brewer(palette = "Pastell1")

sentiments %>%
  mutate(
    avg = (afinn+bing+jockers+loughran+nrc)/5,
    max = pmax(afinn,bing,jockers,loughran,nrc),
    min = pmin(afinn,bing,jockers,loughran,nrc)
  ) %>%
  filter(Company == "KEPPEL CORPORATION LIMITED") %>%
  filter(Year == 2015) %>%
  ggplot(aes(x=page,y=avg)) +
  geom_errorbar(aes(ymin=min,ymax=max),width=0.5,alpha=0.3,
position=position_dodge(0.05)) +
  geom_line() +
  geom_point(aes(color=avg)) +
  theme_minimal() +
  theme(legend.position="none") +
  labs(y = "average Sentiment score") +
  scale_fill_brewer(palette = "RdYlGn")

```

```

# Graphing Sentiment by Year
sentiments_with_average %>%
  mutate(
    Year = as.character(Year)
  ) %>%
  select(Company, Year, page, average_sentiment) %>%
  group_by(Company, Year) %>%
  summarise(average_sentiment = sum(average_sentiment)/
n_distinct(page)) %>%
  ggplot(aes(x=Year,y=average_sentiment)) +
  geom_boxplot(aes(x=Year), fill = "grey", alpha = 0.2) +
  geom_jitter(width = 0.1, alpha = 0.5) +
  theme_minimal() +
  ggtitle("Average Sentiment Score, adjusted for number of pages by
Year") +
  ylab("Average Sentiment Score") +
  xlab("Disclosure for Year Ended In")
  theme(legend.position="none") +
  scale_fill_brewer(palette="Blues")

```

Readability

```

# Add readability scores
for (i in 1:nrow(text_df)){
  flesch_temp <- textstat_readability(
    text_df[i,]$text,
    measure = "Flesch",
    remove_hyphens = TRUE,
    min_sentence_length = 1,
    max_sentence_length = 10000,
    intermediate = FALSE
  )$Flesch
  if (i == 1) {
    flesch_scores <- c(flesch_temp)
  }
  else {
    flesch_scores <- c(flesch_scores, flesch_temp)
  }
}

text_df_with_flesch <- text_df
text_df_with_flesch$flesch <- flesch_scores

# Graph readability
text_df_with_flesch %>%
  select(Company, Year, flesch) %>%
  mutate(
    Year = as.character(Year)
  ) %>%
  ggplot(aes(x=Year,y=flesch)) +
  geom_boxplot(aes(x=Year), fill = "grey", alpha = 0.2) +
  geom_jitter(width = 0.1, alpha = 0.5) +
  theme_minimal() +
  ggtitle("Readability Ease by Year") +
  ylab("Readability Ease") +
  xlab("Disclosure for Year Ended In")
  theme(legend.position="none") +
  scale_fill_brewer(palette="Blues")

```

Text Cleaning (LDA)

```
# Text Cleaning
toSpace <- content_transformer(function (x , pattern ) gsub(pattern,
" ", x))
docs <- tm_map(docs, toSpace, "\n")
docs <- tm_map(docs, toSpace, "/")
docs <- tm_map(docs, toSpace, "@")
docs <- tm_map(docs, toSpace, "\\|")
# docs <- tm_map(docs, toSpace, "*")
docs <- tm_map(docs, toSpace, "[^[:alnum:] ]")

# Convert the text to lower case
docs <- tm_map(docs, content_transformer(tolower))
# Remove numbers
docs <- tm_map(docs, removeNumbers)
# Remove english common stopwords
docs <- tm_map(docs, removeWords, stopwords("english"))
# Remove english (SMART) common stopwords
docs <- tm_map(docs, removeWords, stopwords("SMART"))
## Remove additional stopwords as identified (See Appendix II)
to_remove <- as.vector(read.csv("2. to_remove.csv",header=T)[[1]])
docs <- tm_map(docs, removeWords, to_remove)
# Remove punctuations
docs <- tm_map(docs, removePunctuation)
# Eliminate extra white spaces
docs <- tm_map(docs, stripWhitespace)
# Text lemmization
docs <- lemmatize_words(docs, dictionary = lexicon::hash_lemmas)
```

Topic Modelling (LDA)

```
# Compute Document Term Matrix where Word >= minimumFrequency
minimumFrequency <- 10
DTM <- DocumentTermMatrix(docs, control = list(bounds = list(global
= c(minimumFrequency, Inf))))

# Store rows that are dropped
index_nonzero <- apply(DTM,1,FUN=sum)!=0
DTM <- DTM[index_nonzero,]

# Create models with different number of topics
result <- ldatuning::FindTopicsNumber(
  DTM,
  topics = seq(from = 2, to = 50, by = 1),
  metrics = c("CaoJuan2009", "Deveaud2014"),
  method = "Gibbs",
  control = list(seed = 77),
  verbose = TRUE
)

# Plot evaluation metrics of CaoJuan2009 and Deveaud2014
FindTopicsNumber_plot(result)

# Set number of topics
K <- 7

# Set random number generator seed
set.seed(1234)

# Compute the LDA model, inference via 500 iterations of Gibbs
sampling
topicModel <- LDA(DTM, K, method="Gibbs", control=list(iter = 500,
verbose = 25))

topics <- tidy(topicModel, matrix = "beta")

# Get the top ten terms for each topic
top_terms <- topics %>%
  group_by(topic) %>%
  top_n(10, beta) %>%
  ungroup() %>%
  arrange(topic, -beta)

# Form topics
tmResult <- posterior(topicModel)
beta <- tmResult$terms
theta <- tmResult$topics

terms(topicModel, 10)
exampleTermData <- terms(topicModel, 10)

top5termsPerTopic <- terms(topicModel, 5)
topicNames <- apply(top5termsPerTopic, 2, paste, collapse=" ")

theta_df <- data.frame(theta)
```

```

# Filter for pages which are in the document corpus
index_nonzero <- data.frame(index_nonzero,header=0)$index_nonzero

theta_df$Company <- text_df[index_nonzero,]$Company
theta_df$Year <- text_df[index_nonzero,]$Year
theta_df$page <- text_df[index_nonzero,]$page

theta_df # Distribution of topics

# Re-rank top topic terms for topic names
topicNames <- apply(lda::top.topic.words(beta, 5, by.score = T), 2,
paste, collapse = " ")

# Getting the most probable topics in the entire collection?
topicProportions <- colSums(theta) / nDocs(DTM) # mean probabilities
over all paragraphs
names(topicProportions) <- topicNames # assign the topic names
we created before
sort(topicProportions, decreasing = TRUE) # show summed proportions
in decreased order

# Generating wordcloud
for (i in 1:K){
  ## Topic Wordcloud
  topicToViz <- i
  top100terms <-
    sort(tmResult$terms[topicToViz,],decreasing=TRUE)[1:100]
  words <- names(top100terms)
  # Extract the probabilities of each of the 40 terms
  probabilities <-
    sort(tmResult$terms[topicToViz,],decreasing=TRUE)[1:100]
  # Visualize the terms as wordcloud
  mycolors <- brewer.pal(8, "Dark2")

  # Generate the Word Cloud
  png(paste(i, ".png", sep=""), width=480,height=480)
  wordcloud(words, probabilities,
    random.order = FALSE, color = mycolors, rot.per=0)
  dev.off()
}

```


Regression Analysis

```
# Correlation Matrix
res <- cor(data %>%
  select(-c(Company, Year))
round(res, 2)

library(corrplot)
corrplot(res, type = "upper", order = "hclust",
  tl.col = "black", tl.srt = 45)

# Analysis for ROA
roa_rq2_w <- lm(ROA ~ w_Community + w_Resources + w_Customers +
w_Employees + w_Metholodgy + w_Governance + w_Climate, data )
summary(roa_rq2_w)

roa_rq3 <- lm(ROA ~ flesch + sentiment, data)
summary(roa_rq3)

# Analysis for MV
mv_rq2_w <- lm(MV_t4 ~ BV_t + EARN_t + EARN_t * NEG_t + w_Community
+ w_Resources + w_Customers + w_Employees + w_Metholodgy +
w_Governance + w_Climate, data )
summary(mv_rq2_w)

mv_rq3 <- lm(MV_t4 ~ BV_t + EARN_t + EARN_t * NEG_t + flesch +
sentiment, data)
summary(mv_rq3)

# Analysis for Zmijewski Score
zs_rq2_w <- lm(zmijewski_t ~ w_Community + w_Resources + w_Customers
+ w_Employees + w_Metholodgy + w_Governance + w_Climate, data )
summary(zs_rq2_w)

zs_rq3 <- lm(zmijewski_t ~ flesch + sentiment, data)
summary(zs_rq3)
```

Appendix II: Additional Stop Words Removed

The removal of stop words is done to leave the corpus with only words which are useful to the analysis. In addition to the dictionary of English stop words from Feinerer et al. (2008) and SMART information retrieval system (Lewis et al., 2004), there are other stop words which are removed in this analysis as they are deemed irrelevant. These are presented below, summarised in three categories:

Company-related		Country-related	Irrelevant
singtel uob corporation jardine thaibev keppel wilmar organisation cdl reit ascendas sia airlines sats comfortdelgro ltd uob dbs rws limited	capitaland mapletree beverage sgx sembcorp mapletree beverage cycle carriage venture logistics trust banking yangzijiang developments uol group frasers matheson	thai thailand asean baht indonesia malaysia australia hongkong hong kong singapore australia china world international	words com non www via esg year total page sustainability sustainable related reporting period report fy

Appendix III: Evaluation Metrics of “CaoJuan2009” and “Deveaud2014”

Topics	CaoJuan2009	% Chg	Cum % Chg	Deveaud2014	% Chg	Cum % Chg
2	0.178922			2.558024		
3	0.163802	-8.5%	-8.5%	2.632196	2.9%	2.9%
4	0.143631	-12.3%	-19.7%	2.797818	6.3%	9.4%
5	0.140098	-2.5%	-21.7%	2.818084	0.7%	10.2%
6	0.133208	-4.9%	-25.5%	2.841713	0.8%	11.1%
7	0.125854	-5.5%	-29.7%	2.942365	3.5%	15.0%
8	0.123863	-1.6%	-30.8%	2.90717	-1.2%	13.6%
9	0.120739	-2.5%	-32.5%	2.893421	-0.5%	13.1%
10	0.11088	-8.2%	-38.0%	2.946369	1.8%	15.2%
11	0.101867	-8.1%	-43.1%	2.96754	0.7%	16.0%
12	0.100075	-1.8%	-44.1%	2.980993	0.5%	16.5%
13	0.097093	-3.0%	-45.7%	2.937071	-1.5%	14.8%
14	0.100469	3.5%	-43.8%	2.968951	1.1%	16.1%
15	0.097122	-3.3%	-45.7%	2.967222	-0.1%	16.0%
16	0.088731	-8.6%	-50.4%	2.987292	0.7%	16.8%
17	0.089783	1.2%	-49.8%	3.016476	1.0%	17.9%
18	0.087201	-2.9%	-51.3%	3.004284	-0.4%	17.4%
19	0.087901	0.8%	-50.9%	3.02097	0.6%	18.1%
20	0.082886	-5.7%	-53.7%	3.019461	0.0%	18.0%
21	0.085934	3.7%	-52.0%	3.015959	-0.1%	17.9%
22	0.084143	-2.1%	-53.0%	3.027008	0.4%	18.3%
23	0.081654	-3.0%	-54.4%	3.043813	0.6%	19.0%
24	0.08253	1.1%	-53.9%	3.018318	-0.8%	18.0%
25	0.075137	-9.0%	-58.0%	3.055366	1.2%	19.4%
26	0.073615	-2.0%	-58.9%	3.059574	0.1%	19.6%

27	0.073258	-0.5%	-59.1%	3.05692	-0.1%	19.5%
28	0.07125	-2.7%	-60.2%	3.053202	-0.1%	19.4%
29	0.066505	-6.7%	-62.8%	3.056661	0.1%	19.5%
30	0.071668	7.8%	-59.9%	3.038099	-0.6%	18.8%
31	0.071474	-0.3%	-60.1%	3.050055	0.4%	19.2%
32	0.069479	-2.8%	-61.2%	3.038788	-0.4%	18.8%
33	0.067446	-2.9%	-62.3%	3.026222	-0.4%	18.3%
34	0.069356	2.8%	-61.2%	3.033862	0.3%	18.6%
35	0.067	-3.4%	-62.6%	3.044145	0.3%	19.0%
36	0.065924	-1.6%	-63.2%	3.036874	-0.2%	18.7%
37	0.065304	-0.9%	-63.5%	3.034051	-0.1%	18.6%
38	0.056809	-13.0%	-68.2%	3.077011	1.4%	20.3%
39	0.063551	11.9%	-64.5%	3.041445	-1.2%	18.9%
40	0.059407	-6.5%	-66.8%	3.042395	0.0%	18.9%
41	0.057786	-2.7%	-67.7%	3.06114	0.6%	19.7%
42	0.061882	7.1%	-65.4%	3.015757	-1.5%	17.9%
43	0.062752	1.4%	-64.9%	3.023293	0.2%	18.2%
44	0.06126	-2.4%	-65.8%	3.018354	-0.2%	18.0%
45	0.060804	-0.7%	-66.0%	3.023304	0.2%	18.2%
46	0.05906	-2.9%	-67.0%	3.027811	0.1%	18.4%
47	0.055285	-6.4%	-69.1%	3.038697	0.4%	18.8%
48	0.054731	-1.0%	-69.4%	3.056538	0.6%	19.5%
49	0.052342	-4.4%	-70.7%	3.079457	0.7%	20.4%
50	0.055961	6.9%	-68.7%	3.056176	-0.8%	19.5%

Appendix IV: Readability Scores for Sustainability Reports

Year	Count of Reports	Average RE	Min. RE	Max. RE	Std. Dev. RE
DAIRY FARM INT'L HOLDINGS LTD	7	27.878	22.529	37.747	5.602
YANGZIJIANG SHIPBLDG HLDGS LTD	7	22.763	15.282	28.710	4.800
COMFORTDELGRO CORPORATION LTD	7	16.132	4.947	23.479	5.820
UOL GROUP LIMITED	7	14.133	9.121	21.856	4.314
THAI BEVERAGE PUBLIC CO LTD	7	13.939	10.405	17.026	2.420
SINGTEL	7	12.363	9.502	14.320	1.541
JARDINE MATHESON HLDGS LTD	7	12.149	-6.925	27.771	10.666
CAPITALAND INVESTMENT LIMITED	1	12.024	12.024	12.024	N/A
DBS GROUP HOLDINGS LTD	7	12.023	3.835	26.433	7.167
SINGAPORE AIRLINES LTD	7	11.835	9.268	16.432	2.508
UNITED OVERSEAS BANK LTD	7	10.703	-7.858	27.430	13.550
JARDINE CYCLE & CARRIAGE LTD	7	10.454	4.641	19.304	5.451
WILMAR INTERNATIONAL LIMITED	7	9.986	6.004	13.035	2.837
MAPLETREE COMMERCIAL TRUST	7	9.614	2.640	19.866	7.515
SINGAPORE TECH ENGINEERING LTD	7	9.411	-6.993	17.877	8.957
CAPITALAND INTEGRATED COMM TR	7	9.296	2.681	23.430	7.141

GENTING SINGAPORE LIMITED	7	7.476	-2.250	14.968	6.347
FRASERS LOGISTICS & COMMERCIAL TRUST	5	6.998	-0.577	11.843	4.766
KEPPEL CORPORATION LIMITED	7	6.574	-1.225	9.698	3.564
KEPPEL DC REIT	7	6.188	0.876	10.671	3.455
SATS LTD.	7	4.596	-2.996	18.058	7.101
MAPLETREE INDUSTRIAL TRUST	7	4.442	-7.450	12.221	6.889
CITY DEVELOPMENTS LIMITED	7	4.398	0.623	5.322	1.685
VENTURE CORPORATION LIMITED	5	3.707	-3.436	9.807	5.197
SEMBCORP INDUSTRIES LTD	7	3.519	-5.802	8.639	6.289
OVERSEA-CHINESE BANKING CORP	7	3.443	-15.434	23.832	12.935
ASCENDAS REAL ESTATE INV TRUST	7	3.257	-5.399	13.728	6.063
MAPLETREE LOGISTICS TRUST	7	1.715	-4.972	9.695	5.514
SINGAPORE EXCHANGE LIMITED	7	1.460	-5.715	11.069	5.188
HONGKONG LAND HOLDINGS LIMITED	7	1.093	-11.893	19.482	11.754