Table of Contents

Do management-oriented analysts perform better?

1. Introduction

Analysts play key roles on delivering firm's information to the market and they have varying styles when investigating a firm. Concurrently, as a crucial part in corporate governance, management teams can dominantly influence firm operations and strategies, which will eventually reflect in their firm's value. So, we are curious if analysts can effectively interpret management teams intentions and whether this type of information is prioritized when evaluating a firm's growth potentials.

For this management-oriented information, it may provide analysts with deeper insights into management future plans & strategies, rather than simply pay attention to the financial figures. However, this type of information, being qualitative and subjective (i.e., management-style and preference), may also introduce noise and bias, especially if the management attempts to mislead analysts' judgments. Consequently, it remains uncertain whether analysts who favour management-oriented information can produce superior forecast reports. We refer to this cohort of analysts, which mentioned management related information more frequently in their forecast reports, as management-oriented analysts.

In this proposal, we aim to employ textual analysis techniques to measure whether and the extent to which analysts are management oriented. Our goal is to present descriptive evidence on analysts' varying attentions on managerial information and its consequences. We will use topic modelling to explore what themes management-oriented (MO) analysts usually discuss. Then, comparing the top themes difference between conference call and analyst report. We predict MO analysts can discuss more information that beyond the conference calls discuss. Further, we'd like to know whether major changes in management team can be detected and whether reactions from managers will be different when facing this type of analysts. It may show that management related information could contain key insights that are not able to be seen on financial metrics.

This proposal may examine "black box" of how analysts produce information and what makes a good analyst. It may contribute to the literature of discussing whether management teams possess distinct influence and information that could significantly determine firms' value.

2. Question Background and Literature Review

2.1 Limited Attention, Management Orientation and Management Related Information

When conducting research on firms, analysts may adopt varying areas of focus. According to the behavioral literature, sensory memory will select just specific subsets of the information it receives. This memory system helps in the allocation of cognitive resources. Moreover, it operates as a filter for information received from outside and then transmits selective information to the memory system (Cowan, 1988). When the cognitive load of information surpasses the system's capacity, limited attention comes into play, which consequently means that not all information is processed (Sweller et al., 2012). So, analysts can only have limited attention when they predict one firm's potential since the filter prevents the limited capacity cognitive system from becoming overloaded.

Thus, limited attention will be allocated specific subsets of orientation based on analysts' preference. Some scholars refer that an individual's inherent tendency to exhibit interest and responsiveness towards others is person orientation (Little, 1976; Graziano et al., 2011; McIntyre and Graziano, 2016; McIntyre and Graziano, 2019). McIntyre and Graziano (2016) introduce an example to show what is person orientation. It is an incident in a local community in Indiana that two witnesses of the same event recalled distinctly different aspects when making statements — one focused on the victim, the other on the weapon. Despite the anecdote suggesting person or object orientations are allocated at opposite ends, existing research show that these to be "orthogonal", and measurable independently (Tay et al., 2011; Graziano et al., 2012; Woodcock et al., 2013;). Consequently, we propose that analysts will independently decide how much management related information and how many financial metrics they want to use. These two are orthogonal and are not interact with each other. Since analysts' information preference is determined by cognitive memory systems, it can be seen as innate skills that are randomly assigned across the sample. This proposal emphasizes person orientation and would like to extend this concept into management orientation in accounting and finance area.

As analysts in accounting and finance, person-orientation implies that this subgroup of analyst will mainly focus on the management team. The reason is that person orientation influences an individual's learning pattern, career trajectories, and interpersonal abilities (Graziano et al., 2012; Mayer and Skimmyhorn, 2017; Bairaktarova and Pilotte, 2020; McIntyre et al., 2021). This person orientation will guide analysts' interests and attention toward people when they produce reports. As a result, their thoughts and learning motivations will unconsciously lead them to pay attention to managers in firms where they follow and give them evaluations. So, their reports will spontaneously include some markers and efforts that reflect management-related information (Mayer and Skimmyhorn, 2017; McIntyre et al., 2021). We define management-oriented analysts as individuals who focus on management traits and tend to learn more about management intended actions from site visits, conference calls, proxy statements, etc. Meanwhile, management related information is not constrained by managers themselves, it also covers managerial qualitative information conveyed by the management teams like firm strategies, research directions, Green House Gas (GHG) Emission, Environmental, Social and Governance (ESG) ambitions and so on.

In this proposal, we will try to apply textual analysis methods into analysts' research reports to construct measures of representing if they are management oriented. We will also explore what managerial topics management-oriented analysts usually discuss and whether they can detect management major changes (i.e., CEO turnover) in advance. Based on the discussion above, we consider what the measurement of management orientation is. To capture it, prior literature indicates that individuals' choice of what to write mirrors their orientation of cognitive system: toward person or not (Kahneman, 1973; Mole, 2013; McIntyre and Graziano, 2016). Above all, we argue that management-oriented analysts will mention the manager's name more frequently and tend to ask more manager's qualitative information (i.e., teams change; early experience; background; previous career; their ambitions and plans about the firm;) in the conference calls. Subsequently, manager's avoid-repsonse rates in the conference calls will decrease when the question is brought by management-oriented analysts. The reasoning process is that nondisclosure from the specific question that focuses on the individual is a negative signal towards investors (Grossman, 1981), and thereby will easily increase the media coverage (attention) and do harm to manager's reputation (Zhou and Zhou, 2020).

In this research, we probe the information gathering strategies of management-oriented analysts by comparing the nature of questions they pose during earnings calls with those raised by analysts who are not management-oriented. By delving into the research reports of management-oriented analysts and scrutinizing their methods of information collection during earnings calls, our objective is to amplify our understanding of the analyst's role in learning qualitative information directly from management.

2.2 Textual Analysis in Accounting and Previous Application

In this proposal, we plan to employ techniques such as counting occurrences of manager names, identifying topics in the reports, and detecting unanswered questions or significant changes, among others. These methods will aid in exploring our central question: Do management-oriented analysts outperform their non-management-oriented counterparts?

2.2.1 Simple Transformation

Textual analysis techniques have been employed by accounting scholars for over a decade. These methods have offered considerable insights into various aspects of accounting and financial research. Simple Transformation is widely used in textual analysis in accounting. It uses some traits in the textual data (i.e., number counts) as inputs and then transfer into the models to produce outputs with minimal transformation. Butler et al. (2004) firstly employ the textual analysis to conduct keyword search to explore auditing opinions in auditing reports. Following the first accounting study in textual analysis, Li et al. (2008) construct a measure to capture reports' readability, and other scholars use Simple Transformation to form measures of sentiment, disclosure quantity, and forward-looking keywords (e.g., Kothari et al. 2009; Frankel et al. 2010; Li 2010).

However, explaining the results of these outputs usually face some questions. We need to justify why we can use the measure mentioned in the paper to proxy the key variable of research question. As a result, interpretation often relies on intuition, behavioural science, psychology, among other fields of research. Therefore, when employing Simple Transformation, it's crucial that we clearly articulate the rationale behind the construction of our proxy measures. The logic behind our metrics should be firmly grounded in relevant theory and empirical findings, enabling more accurate interpretation of the results.

2.2.2 Machine Learning and Deep Learning Application: Topic Modelling

With the widespread adoption of Simple Transformation in accounting research, methodologies involving Machine Learning (ML) and Deep Learning (DL) have become increasingly popular. Among these, topic discovery often combines with ML or DL in the accounting field. Topic discovery aims to summarize a document by clustering words into themes known as topics. A topic model uncovers a set of latent topics and determines their distribution in a specific document.

For Topic Modelling, Latent Dirichlet Allocation (LDA) is the most prevalent unsupervised Natural Language Processing (NLP) model for theme analysis. Originally, LDA was designed to work with individual word counts within the Bag of Words (BOW) text representation approach. Campbell et al. (2014) were the first to employ LDA in accounting. They utilized the model to identify frequently occurring words related to risk disclosure and construct a new way to evaluate the quality and quantity of risk index. Some other scholars continue using LDA to compare and observe the topic dynamics of reports produced by the financial markets. They are usually interested in comparing the firms conference calls' topics with analyst reports' topics to explore the information role of analysts, trends of core topics evolving in 10-K filings, and fraud detection using textual traits besides the financial metrics (Dyer et al., 2017; Huang et al., 2018; Brown et al., 2020).

However, LDA is time-consuming and resource-intensive due to the large vocabulary created by BOW. At the meantime, LDA is an unsupervised model that doesn't use pre-labelled examples to learn, and thus, often creates topics that are not necessarily helpful or clear to the researchers. In addition, it doesn't guarantee the discovery of specific topics researchers might be interested in.

To manage the problem of high-dimensionality, Dieng et al. (2020) introduced the Embedded Topic Model (ETM) which uses word embeddings as inputs for Latent Dirichlet Allocation (LDA). Word embeddings, like Word2Vec, efficiently decrease the high-dimensional data, particularly for infrequent words, while simultaneously capturing the semantic relationships among words. This development offers fresh insights and potential for textual analysis in the domain of accounting research. It also hints us that we need to know if topics uncovered by the model could be interpretable economically.

2.2.3 Named Entity Recognition in Finance and Accounting

Named Entities are things existed in the document that reflect real-world entity. Name Entity Recognition (NER) is to identify our interests of entities such as location, organization and individual's name. Using machine learning, NER not only are able to find out the exact text in the training, but also can generalize in new dataset.

In Computer Science, lots of pre-trained NLP model can conduct NER process (Nadeau and Sekine, 2007). These models can recognize named entities, including geographic locations. However, based on anecdotal experience (informal, personal observations), these models often make a substantial number of both Type I and Type II errors (Mansouri et al., 2008) particularly when they are applied to corporate filings (A Type I error, also known as a false positive, happens when the model incorrectly identifies a named entity. A Type II error, also known as a false negative, occurs when the model fails to identify a named entity that is present).

So, these pre-trained models, while convenient, may not be sufficiently accurate "out of the box" for certain tasks, such as analysing corporate filings in accounting and finance. This means that they may require further fine-tuning - adjusting the model's parameters and training it on a more specific task or dataset - to improve their performance.

Recently, Shah et al. (2023) have trained a NER model specified in finance related research (FiNER) and published it in arXiv. They assert it has high quality when identify the finance filing and business-related entities. This FiNER Model's weak-ner framework and dataset can be found in hugging face and GitHub. Even though it is promising, their model needs to be examined by peers in finance field.

3. Research Design

3.1 Data Collection and Processing

For sample range, we use Standard & Poor's 500 (S&P 500) firms as they are worthwhile for analysts to investigate and are the mainstream for analysts to gain benefits in the market. We will extract names from brokerage reports which can be found in Investext. We first transform PDF into TXT format. Then, filter and remove useless descriptions in reports such as tables, standard boilerplate sentences, no textual information reports, and analyst contacts. After that, we match analyst names from header part extracted by Investext with I/B/E/S analyst names. So, we can have analyst-report level information.

```python
import PyPDF2
import pandas as pd

# Function to convert PDF file to TXT
def convert_pdf_to_txt(file_path):
    pdf_file = open(file_path, 'rb')
    pdf_reader = PyPDF2.PdfFileReader(pdf_file)
    text = ''
    for page_num in range(pdf_reader.numPages):
        text += pdf_reader.getPage(page_num).extractText()
    pdf_file.close()
    return text

# Read analyst names from Investext and I/B/E/S
investext_analyst_names = pd.read_csv('investext_analyst_names.csv')
ibes_analyst_names = pd.read_csv('ibes_analyst_names.csv')

# Iterate through the PDF files
for file in pdf_files:
    text = convert_pdf_to_txt(Users/eric/Desktop/学业/)

# Match analyst names between Investext and I/B/E/S
matched_analysts = pd.merge(investext_analyst_names, ibes_analyst_names, on='analyst_name')
```

```python
# Remove boilerplate sentences
    text = re.sub(r'This report is intended for informational purposes only.', '', text)

# Remove tables (they are structured with lines starting with 'Table')
    text = re.sub(r'\n\Table.*', '', text)

# Remove contact information
    text = re.sub(r'For further information, please contact.*', '', text)
```

3.2 Construct Measure of Management-Oriented (MO) Analysts

We then use Stanford NER (with Conditional Random Field (CRF) model) to identify if each report contains management individual names. To learn about the suitability of choosing this technique in financial text, please see Appendix.

The algorithm is designed to identify names and categorize them in pre-specified group. It can also extract numerical expressions, like those indicating amount of money or percentage in some financial metrics. We feed the discussion sections of analyst reports into the CRF classifier in order to extract named entities of the 'person' type. We expect that analysts who tend to focus more on management team will mention management team names more frequently in their report discussions. Then, we can know in what percentage one analyst mentions the management names across his/her reports. We hypothesis that the percentage of reports one analyst mentions management names more than the median percentage of management-mentioned reports for each analyst will be the standard to define management-oriented analyst.

```
import stanfordnlp

stanfordnlp.download('en')  # Download the English models

nlp = stanfordnlp.Pipeline(processors='tokenize,ner')  # Initialize the pipeline

def check_for_names_in_text(text):
    doc = nlp(text)  # Process the text
    for sentence in doc.sentences:
        for token in sentence.tokens:
            if token.ner != 'O':  # 'O' means no named entity
                if 'PERSON' in token.ner:  # Check if the entity is a person

                    print('Name found:', token.text)

report = “/Users/eric/Desktop/学业/”
check_for_names_in_text(report)
```

3.3 MO Analysts May Do Better in Information Gathering and Processing

3.3.1 Incremental Topics between Analyst Reports and Conference Calls

Now that we have analyst reports and their orientations, we will try to get the conference call meeting minutes from Thomson Reuter’s StreetEvents. Since each industry has specific topics, we will conduct LDA grouped by each industry. In accounting and finance, using four-digit Standard & Poor’s Global Industry Classification Standard (GICS) to identify industries is better than using SIC codes. They are good at classify which industry the firm should be (Kadan et al., 2012). We will explore incremental changes from manager’s scripts(meeting minutes) to analyst’s scripts, then compare difference in incremental changes between MO and non-MO analysts. We hypothesis MO analysts can generate more new topics or new words in each topic from manager’s conference calls. LDA can give us the distribution of topic in documents and the distribution of words in the topic, which is useful.

Following Ramage et al. (2009), we will try to use LDA from Stanford Topic Modeling Toolbox. We try not to not to use "stemming" (changing words to their base form) because it's too harsh for financial text (Huang et al., 2018). Here, words with the same base often don't mean the same thing. For example, "marketing" is different from "market"; “accounting” is the special term for the profession and does not have same meaning when refers to “account” (Porter, 1980).

Also, we need to set three parameters for LDA. These include the total number of topics present in the complete collection of documents, as well as parameters α and β. These latter parameters govern the smoothness of the topic and word distributions. According to the Rosen-Zvi et al. (2004), we can use “perplexity score” to reduce uninterpretable topics LDA generated. With more topics are generated, this score will be decreased. But its decrease will diminish gradually. We can find the “turning point” as Ball et al. (2015) discussed. As for α and β, we set 0.1 and 0.01 respectively instructed by Kaplan and Vakili (2015). We expect to see new words in the specific topic compared to manager’s scripts.

```python
import stanfordnlp


# Set your parameters
num_topics = 20  # number of topics will be tested by perplexity score
alpha = 0.1
beta = 0.01

documents = df['report'].tolist()
dictionary = Dictionary(documents)
corpus = [dictionary.doc2bow(document) for document in documents]

# Initialize the LDA model
lda_model = stanfordnlp.Models.LDA(num_topics, alpha, beta)

# Train the LDA model with the reports
for industry, reports in grouped_reports.items():
    lda_model.train(corpus)

#  use it to extract topics from a new document
for industry, conference_call in conference_call_minutes.items():
    topics = lda_model.extract_topics(conference_call)
    # analyze the topics, look for new words compared to manager's scripts, etc.

#  compute the perplexity score of the model
perplexity = lda_model.compute_perplexity()
```

3.3.2 Detect Statement of Management Major Change (MMC) in Analyst’s Report

In the conference calls, if analysts ask about management turnover things, we can find it on Q&A session. But we want to know whether analyst process this management-personnel information. So, we try to examine analysts report for comments on MMC. Using regular expression, we can see whether reports contain “CEO turnover” sentence. Then, we would like to know the percentage of MMC reports in every analyst’s documents and explore the difference between MO and non-MO analysts.

We propose MO analyst will have higher percentage of MMC reports in one analyst’s all reports than non-MO analyst will do. The comments support our prediction MMC is noticed by analysts and is processed into their equity research consideration.

3.3.3 Manager’s Avoid-Response Measurement

We will also use key sentence detection to find the Avoid-Response (AR) answers in all conference scripts. To set the list of AR, we follow the advice from experimental behavioral literature (Okan et al., 2016; Huber and Huber, 2019; Douglas et al., 2023): We could use CloudResearch and Prolific’s participants to annotate the Avoid-Response in Q&A section in all scripts. Subsequently, we can hire research assistants to double check our corpus.

To develop the list, we will use training set to enhance in-sample accuracy and exploit test set to validate the out-sample performance. After that, we can get the list that can be set in our regular expression.

We hypothesize a lower AR rate within the MO analyst group. This supposition is grounded in the belief that the dissemination of personal conversational elements is typically more straightforward than sharing pivotal company metrics with analysts. Nevertheless, personal life events such as impending parenthood or relocation intentions could serve as potential indicators of substantial managerial changes. Also, media coverage will tend to focus more on personal information and managers avoid-response will incur negative signal to the market.

3.4 Minors in Design

The proxy for an MO analyst might not truly reflect their actual preference. It's not immediately clear why names are seen as significant linguistic markers. Also, limited attention could not only differentiate MO and non-MO analysts, but it might also affect their learning abilities and different social connections. This suggests that being Management-Oriented is not necessarily an innate preference, but rather it could be influenced by various factors. How to tease these out needs to be discussed more later.

Appendix

Traces of Textual Analysis in Accounting

"At that time (2000), … the broader accounting academic community had little appreciation on textual analysis … did not feel that my work had a place …desk rejection letters from major accounting journals..."

—Ingrid E. Fisher
University at Albany, SUNY

Accounting records lots of traces of firm's activities, which contains many scripts. Also, almost business activities need communication. So, natural language is the key among business. For me, it means textual analysis can be used in accounting academy. Textual analysis, as an application of Natural Language Processing (NLP), can be used for information capturing and extracting features of scripts. This provides a new sight besides traditional accounting figure.

In the 2017 meeting of the American Economic Association, Shiller (2017) underscored the critical role narratives play. However, it is the power of textual analysis that enables us to comprehend these narratives fully. In essence, textual analysis operates under the assumption that words or word sequences within a text constitute meaningful entities, thereby offering valuable insights upon analysis. In contrast to prior periods, textual analysis is increasingly emerging as a significant avenue within the accounting research landscape.

LDA Topic Model

Latent Dirichlet Allocation (LDA) operates under the premise that the generation of a document follows a two-step process. Initially, a topic is randomly selected according to the specific topic distribution within the document. Subsequently, a word is picked up at random in alignment with the word distribution pertaining to the topic that was previously identified. The iterative execution of these two steps on a word-by-word basis results in the creation of a comprehensive document. This fundamental concept underlying LDA mirrors the procedure that a human writer may employ in the construction of a document: They decide what topics they want to talk about, then they pick words that explain those topics.

Stanford NER

Named Entity Recognition (NER) is a way to find and sort names in text into fixed groups. A good example of this is the Stanford NER. It uses a model called the Conditional Random Field (CRF), early discussed by Lafferty et al. in 2001 and further detailed in financial reports by Hope et al. in 2016. The Stanford NER is very good at predicting categories of names, like PERSON, ORGANIZATION, and LOCATION. It can also find names with almost the same accuracy as a person. This makes it a helpful tool when analyzing financial reports.

Reference

Bairaktarova, D. N., and M. K. Pilotte. 2020. Person or thing oriented: A comparative study of individual differences of first-year engineering students and practitioners. Journal of Engineering Education 109 (2): 230–242.

Ball C, Hoberg G, Maksimovic V 2015. Disclosure, business change and earnings quality. Working paper, University of Maryland, College Park.

Brown, N. C., R. M. Crowley, and W. B. Elliott. 2020. What are you saying? Using topic to detect financial misreporting. Journal of Accounting Research 58 (1): 237–91.

Butler, M., A. J. Leone, and M. Willenborg. 2004. An empirical analysis of auditor reporting and its association with abnormal accruals. Journal of Accounting and Economics 37 (2): 139–65.

Cowan, N., 1988. Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. Psychological Bulletin 104, 163–191.

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

Douglas B D, Ewell P J, Brauer M. Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA[J]. Plos one, 2023, 18(3): e0279720

Dyer, T., M. Lang, and L. Stice-Lawrence. 2017. The evolution of 10-K textual disclosure: Evidence from Latent Dirichlet Allocation. Journal of Accounting and Economics 64 (2): 221–45.

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

Graziano, W. G., M. M. Habashi, and A. Woodcock. 2011. Exploring and measuring differences in person-thing orientations. Personality and Individual Differences 51 (1): 28–33.

Graziano, W. G., M. M. Habashi, D. Evangelou, and I. Ngambeki. 2012. Orientations and motivations: Are you a “people person,” a “thing person,” or both? Motivation and Emotion 36 (4): 465–477.

Grossman, S. J. 1981. “The Informational Role of Warranties and Private Disclosure About Product Quality.” Journal of Law and Economics 24: 461–83.
Hope, O. K., D. Hu, and H. Lu. 2016. The benefits of specific risk-factor disclosures. Review of Accounting Studies 21 (4): 1005–1045.

Huang, A. H., R. Lehavy, A. Y. Zang, and R. Zheng. 2018. Analyst information discovery and interpretation roles: A topic modeling approach. Management Science 64 (6): 2833–55.

Huber, C., Huber, J., 2019. Scale matters: risk perception, return expectations, and invest- ment propensity under different scalings. Experimental Economics 22, 76–100.

Kadan O, Madureira L, Wang R, Zach T 2012. Analysts’ industry expertise. J. Accounting Econom. 54(2–3):95–120.

Kahneman, D. 1973. Attention and Effort. Englewood Cliffs, NJ: Prentice Hall.

Kaplan S, Vakili K 2015. The double-edged sword of recombina- tion in breakthrough innovation. Strategic Management J. 36(10): 1435–1457.

Kothari, S. P., X. Li, and J. E. Short. 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): 1639–70.

Lafferty, J., A. McCallum, and F. Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In Proceedings of the 18th international conference on machine learning, 282–289. Morgan Kaufmann, San Francisco, CA.

Li, F. 2008. Annual report readability, current earnings, and earnings persistence. Journal of Accounting and Economics 45 (2–3): 221–47.

Li, F. 2010. Textual analysis of corporate disclosures: A survey of the literature. Journal of Accounting Literature 29: 143–65.

Little, 1976. Specialization and the varieties of environmental experience: empirical studies within the personality paradigm. In Experiencing the environment, edited by S. Wapner, S. B. Cohen, and B. Kaplan, 81–116. New York: Plenum Press.

Mansouri, A., Affendey, L. S., & Mamat, A. 2008. Named entity recognition approaches. International Journal of Computer Science and Network Security, 8(2), 339-344.

Mayer, J. D., and W. Skimmyhorn. 2017. Personality attributes that predict cadet performance at West Point. Journal of Research in Personality 66: 14–26.

McIntyre, M. M., and W. G. Graziano. 2016. Seeing People, Seeing Things: Individual Differences in Selective Attention. Personality and Social Psychology Bulletin 42 (9): 1258–1271.

McIntyre, M. M., and W. G. Graziano. 2019. A snapshot of person and thing orientations: How individual differences in interest manifest in everyday life. Personality and Individual Differences 136: 160–165.

McIntyre, M., J. L. Gundlach, and W. G. Graziano. 2021. Liking guides learning: the role of interest in memory for STEM topics. Learning and Individual Diffferences 85: 101960.

Mole, C. 2013. Attention is cognitive unison: an essay in philosophical psychology. Oxford, UK:

Nadeau, D., & Sekine, S. 2007. A survey of named entity recognition and classification. Lingvisticae Investigationes, 30(1), 3-26

Okan, Y., Galesic, M., Garcia-Retamer, R., 2016. How people with low and high graph literacy process health graphs: Evidence from eye-tracking. Journal of Behavioral Decision Making 29, 271–294.

Porter MF (1980) An algorithm for suffix stripping. Program, 14(3):130-137

Ramage D, Rosen E, Chuang J, Manning C, McFarland DA (2009) Topic modeling for the social sciences. In NIPS Workshop on Applications for Topic Models: Text and Beyond.

Rosen-Zvi M, Griffiths T, Steyvers M, and Smyth P (2004) The author-topic model for authors and documents. In Proc. of the 20th conf. on Uncertainty in artificial intelligence. (UAI, Banff, Canada), 487- 494

Shiller, R. J. 2017. Narrative Economics. American Economic Review 107 (4): 967–1004.

Song, S. 2021. The informational value of segment data disaggregated by underlying industry: Evidence from textual features of business descriptions. The Accounting Review 96 (6): 361-396.

Sweller, J., Ayers, P., Slava, K., 2012. Cognitive load theory. Springer.

Tay, L., R. Su, and J. Rounds. 2011. People-things and data-ideas: Bipolar dimensions? Journal of Counseling Psychology 58 (3): 424–440.

Woodcock, A., W. G. Graziano, S. E. Branch, M. M. Habashi, I. Ngambeki, and D. Evangelou. 2013. Person and Thing Orientations: Psychological Correlates and Predictive Utility. Social Psychological and Personality Science 4 (1): 116–123.

Zhou, F. S., & Zhou, Y. (2020). The dog that did not bark: Limited price efficiency and strategic nondisclosure. Journal of Accounting Research, 58(1), 155-197.