Article:
Jackson, S., Vanteeva, N. and Fearon, C. (2019) An investigation of the impact of data breach severity on the readability of mandatory data breach notification letters: evidence from U.S. firms. Journal of the Association for Information Science and Technology, 70 (11). pp. 1277-1289. ISSN 2330-1635

https://doi.org/10.1002/asi.24188

An Investigation of the Impact of Data Breach Severity on the Readability of Mandatory Data Breach Notification Letters: Evidence from U.S. Firms

Stephen Jackson*
School of Management, Royal Holloway, University of London, Egham, TW20 0EX, UK.
Email: Stephen.Jackson@rhul.ac.uk.

Nadia Vanteeva
Department of Economics, University of the West of England, Bristol, BS16 1QY, UK.
Email: Nadia.Vanteeva@uwe.ac.uk.

Colm Fearon
Christ Church Business School, Canterbury Christ Church University, Canterbury, CT1 1QU, UK.
Email: colm.fearon@canterbury.ac.uk.

**corresponding author*

Abstract

The aim of this paper is to investigate the impact of data breach severity on the readability of mandatory data breach notification letters. Using a content analysis approach to determine data breach severity attributes (measured by the total number of breached records, type of data accessed, the source of the data breach and how the data was used), in conjunction with readability measures (reading complexity, numerical intensity, length of letter, word size and unique words), 512 data breach incidents from 281 U.S. firms across the 2012 – 2015 period are examined. The results indicate that data breach severity has a positive impact on reading complexity, length of letter, word size and unique words, and a negative impact on numerical terms. Interpreting the results collectively through the lens of impression management, it can be inferred that business managers may be attempting to obfuscate bad news associated with high data breach severity incidents by manipulating syntactical features of the data breach notification letters in a way which makes the message difficult for individuals to comprehend. The paper contributes to the information studies and impression management behavior literatures, by analyzing linguistic cues in notifications following a data breach incident.

Introduction

Data breach, defined as a security incident whereby confidential, safeguarded or sensitive data[1] is copied, observed, stolen, transferred or used by an unauthorized person represents a growing risk for society and organizations (Sen & Borle, 2015). The World Economic Forum list massive incidents of data fraud or theft as one of the top five global risks in terms of likelihood for 2018 (World Economic Forum, 2018). Moreover, the annual number of data breaches in the U.S. grew from 1,093 million in 2016 to 1,579 million in 2017 (Statista, 2018). Consequently, data breaches can have a huge impact on the organization and individuals affected. The Ponemon Institute found the total cost of a breach for an organization to be $3.86 million, with data breaches being the costliest in the U.S. (Ponemon Institute, 2018).

These challenges are further exacerbated by the growing requirement for organizations to notify consumers that a data breach involving personal information has occurred. In North America, all 50 states have enacted laws requiring data breach notification (NCSL, 2018). Data breach notification can be understood as a form of communication, usually in writing, to alert affected consumers of the potential risks associated with a data breach event, as well as the appropriate action which can be taken. Notwithstanding the importance of letters as a popular data breach notification choice, with many governments offering standardized templates and regulatory advice in relation to structure and content, it is important to stress that firms can still control many features of the letter (Veltsos, 2012). Since data breach notification can be perceived as a form of negative news which can have legal, financial and reputational consequences, we suggest that business managers may have incentives to engage in impression management techniques.

In the case of data breach notification, the idea that managers may be manipulating the content of letters in a way which safeguards their own interests is gaining increased traction (Harris, 2014; Narisi, 2012; Ponemon Institute, 2014), but still in need of further empirical support and testing. An important point raised is that, despite managers in many cases being aware of the issue, data breach notification letters are often written in a way that are too complex and thus lack transparency. As noted by Kamala Harris (Attorney General of the California Department of Justice) "*our recommendation to companies and agencies to improve the readability of their breach notices does not seem to have been heeded*". In addition, the reading level of notices submitted in 2012 and 2013 were "*significantly beyond the average reading level of the American population*" (Harris, 2014). Acknowledging that many firms are aware of the issue of reading complexity and are not acting on it, and with much advice acknowledging that data breach notifications should be written in a clear and straightforward manner (Veltsos, 2012), this may be signaling that managers are attempting to obfuscate negative news by manipulating data breach notification characteristics in a way which leaves it more difficult for many American consumers to comprehend.

Given the increased legal requirement for data breach notification to consumers introduced by a growing number of countries, as well as claims that impression management strategies may be more prevalent than first thought, this paper is motivated by the need to examine the notification tactics (reading ease manipulation) used by business managers when disclosing data breach incidents. The paper contributes to the information studies literature on detecting impression management, also referred to as self-presentation, behavior of companies by analyzing linguistic cues in notification letters following a data breach incident. More specifically, while impression management can manifest itself in many forms, one area which the concept has been increasingly applied to in information studies is computer-mediated communication, particularly the online world e.g., blogs, dating sites, online discussion forums and webpages. Studies, for instance, have

[1] Confidential, safeguarded or sensitive data may include personal account data (e.g., user names, email addresses and passwords); financial data (e.g., bank account credentials, credit card data); personal identity information (e.g., social security numbers, date of birth, names, ID numbers, personal medical records); as well as existential data (data of national security importance, trade secrets, intellectual property).

examined the processes/strategies of self-presentation for achieving on-line dating success (e.g., Ellison, Heino & Gibbs, 2006). An individual's personality characteristics (e.g. extraversion, self-esteem, personality traits, individual qualities) can influence impression management tactics in the context of social networking (Kramer & Winter, 2008; Rosenberg & Egbert, 2011). There is also evidence to suggest that: (a) the perceived usefulness and effectiveness of social networking sites are related to self-presentation (Min & Kim, 2015); (b) expressions of self-presentation in the interaction between "asker" and "answerers" on Q&A websites are important (Raban, 2009), and (c) factors (e.g., age, sex, genre, identifiability) may influence the self-motivating strategies of online bloggers (Fullwood, Melrose, Morris & Floyd, 2012). Notwithstanding the importance of this body of research, to the best of our knowledge, from the lens of impression management, no studies have examined the impact of data breach severity on the readability of mandatory data breach notification attributes.

One important question is why do we need to study the impact of data breach severity on the readability of mandatory data breach notification letters? While research has investigated the reading ease of corporate communication, particularly annual reports, much of the focus is on the relationship between readability and corporate financial performance (Merkl-Davies & Brennan, 2007), with few readability studies considering the concept of severity. Since higher data breach severity is more likely to have a negative impact on the firm (Goel & Shawky, 2009), as well as the tendency to downplay the severity of a negative event to prevent reputational harm (Zaharopoulos & Kwok, 2017), it may be the case that when faced with higher levels of data breach severity, business managers may partake in strategies which reflects their self-serving motives when responding to the crisis situation. An examination of formal (written) mandatory notifications is important given that this approach is increasingly perceived as being best practice among other parts of the world as an effective way of communicating the facts associated with data breach incidents, allowing consumers to protect themselves both financially and physically (Ponemon Institute, 2014). Mandatory, as opposed to voluntary (data breach) notification represents an interesting area of study. Managers may attempt to protect themselves, by making involuntary/mandatory letters more difficult to read, especially, if confronted with a high data breach severity level.

Drawing on 512 data breach incidents from 281 North American firms from 2012 -2015, and controlling for several factors – firm size, age of firm, data breach frequency, ownership type, industry, time period and U.S. state, the findings indicate that data breach severity (measured by the total number of breached records, type of data accessed, the source of the data breach and how the data was used) has a positive impact on reading complexity, length of letter, word size and unique words, and a negative impact on numerical terms. Interpreting the results collectively through the lens of impression management, particularly reading ease manipulation (Merkl-Davies & Brennan, 2007), our results and analysis suggest that when faced with higher data breach severity, business managers may be using obfuscation tactics (making the communication more difficult to comprehend) to manipulate notification attributes as a mechanism to protect face.

The remainder of this paper is organized as follows. First, literature relating to this study is briefly outlined, and this is followed by hypotheses. Next, the data and methodology used for this study are described. This is followed by a discussion of the empirical results and concluding remarks.

Literature Review

Impression Management

Impression management, also known as self-presentation, can be referred to as a conscious or unconscious process whereby individuals, both verbally and non-verbally, try to sway the

perceptions of others about an individual, occasion or object. They do this in order to be favorably perceived by others, or to present themselves in the best possible light (Merkl-Davies, 2007). Rather than being grounded in one academic discipline, impression management thinking draws on a number of interdisciplinary areas, including economics, political science, psychology and sociology. Some of the dominant theories influencing impression management, especially how managers manipulate business narrative documents include: agency theory; legitimacy theory; signaling theory; stakeholder theory and institutional theory.

Agency theory seeks to illuminate the relationship between principal (e.g., stakeholders) and agents (e.g., managers). As managers typically discharge their fiduciary duties in an increasingly complex milieu; whereby, financial incentives e.g., salary, promotion, bonuses and risk are all closely tied to business performance, they might become tempted to engage in reporting practices that consider their own interests first. This could be viewed as a way of protecting their personal interests, along with those of management and their shareholders more generally (Abrahamson & Park, 1994). Alternatively, drawing on the assumption that firms are also influenced by the norms of society, legitimacy theory recognizes that in order for the activities of a firm to be perceived as legitimate (trustworthy, desirable, appropriate), it must be seen to be operating within the boundaries of society in which it functions (O'Donovan, 2002). If misalignment (legitimacy gap) occurs between the actions of a firm and how society perceive the organization should behave, for instance, perhaps in mishandling data or dealing with a crisis situation, this can weaken legitimacy which, in turn, can lead to loss of customers, reduced demand for products/services or possibly government levies. As a way to restore or prevent loss of legitimacy, managers may engage in impression management techniques. Accounting for the behavior between two parties (e.g., firms or individuals) with different levels of access to information, signaling theory recognizes that the sender of information must decide if and how to signal (communicate) particular information, and the receiver needs to choose how it plans to act on the information received (Connelly, Hoskisson, Tihanyi & Certo, 2010). For example, when faced with superior performance, firms may act in a manner whereby they attempt to signal this superiority by being more transparent in how they present and disclose this information (Smith & Taffler, 1992).

Stakeholder theory stresses that organizations have a moral responsibility to satisfy the needs of its stakeholders. Since stakeholders (e.g., community, creditors, customers, employees, governmental parties, investors, suppliers and trade unions) have a major interest in the firm and can influence, and be influenced, by the actions of an organization, it is important to respond to the expectations of various stakeholder groups. In order to meet the needs of stakeholders, perhaps in response to various stakeholder requests, organizations may engage in corporate communication practices which seek to satisfy the desires or emulate the values of a particular stakeholder grouping (Hooghiemstra, 2000). Institutional theory recognizes that firms can be swayed and constrained by the institutional environment they belong to. Institutional structures and governance, such as, norms, routines, and schemes, can act as blueprints for forms of social behavior. Conforming to social behavior within the institutional environment is seen as a means to influence the perceptions of others and can enhance organizational support (Pfeffer & Salancik, 1978). As poor adherence to institutional structure and good governance runs the risk of endangering the success of the organization, business managers may write in a biased way which responds and deals with institutional pressures. While these various theories may differ in relation to their origin, focus and method of analysis, an underlying, but central, theme arising is the tendency for individuals to engage in forms of impression management to create a more favorable environment.

Impression Management Behavior/Strategies

Merkl-Davies and Brennan (2007) acknowledge that managers ("preparer perspective") can engage in one of two forms of impression management behavior (1) concealment and (2) attribution.

Concealment occurs when managers attempt to either obfuscate negative news or accentuate good news. Attribution refers to a self-protecting tactic which attempts to shift the blame by transferring responsibility for negative outcomes to another. The authors also outline seven different managerial impression management strategies associated with corporate narrative documents (see figure 1). As it would be unworkable to provide a comprehensive overview and empirically examine each of these strategies in detail, the focus of the study is concealment. We decide not to focus on attribution as the emphasis is often on discretionary disclosure of additional information (performance attributions) to enhance firm reputation (Baginski, Hassell & Hillison, 2000). Furthermore, as data breach notification letters can be viewed as a form of negative news (Veltsos, 2012), the concentration was on "obfuscation of bad news". Given that readability of data breach notification letters is a central concern as raised earlier and may be conducive to impression management, the effort of this paper is on readability (reading ease manipulation). A summary of the characteristics of managerial impression management behavior, together with the position adopted in this paper (highlighted in bold) is illustrated in figure 1.

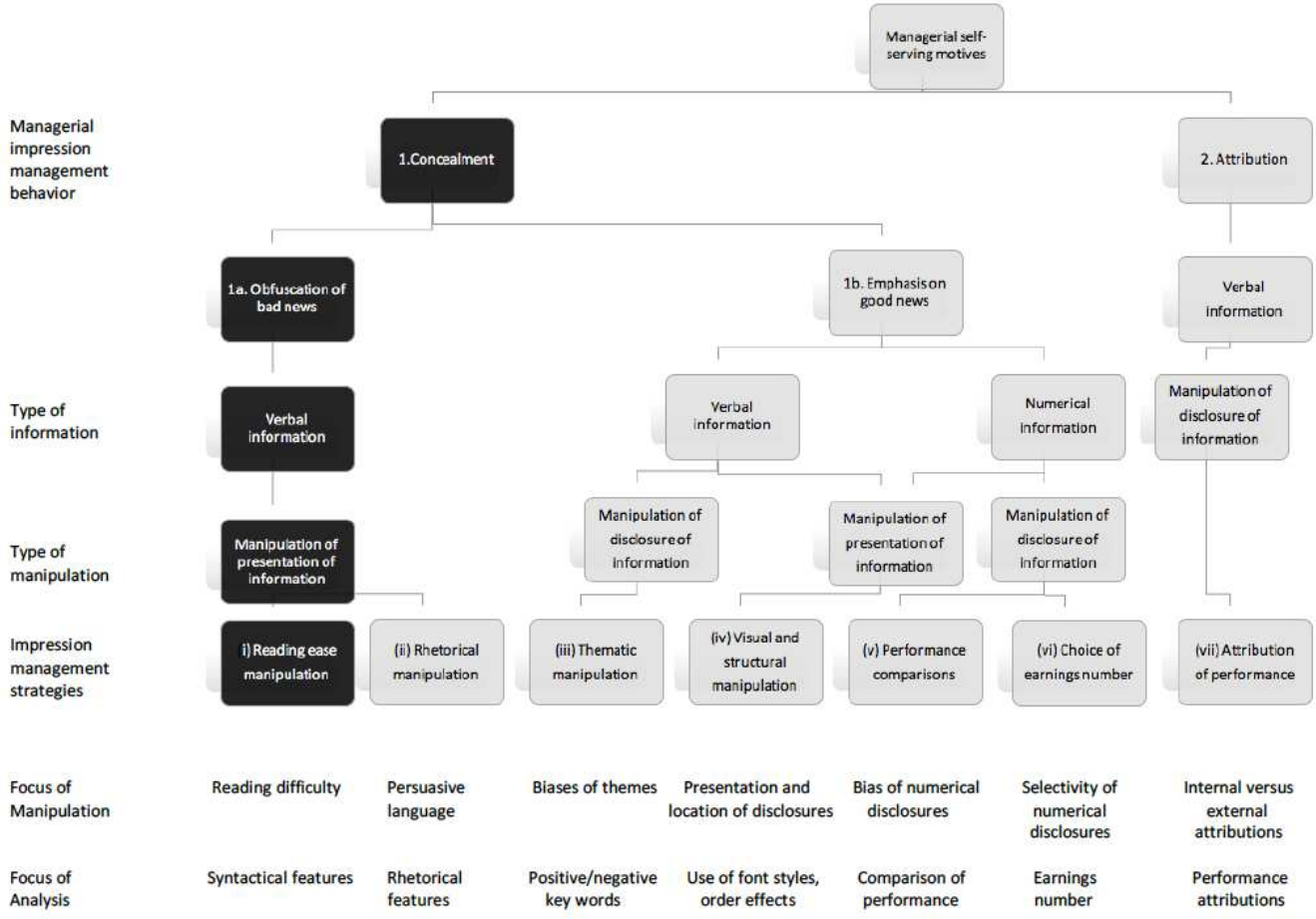

FIG. 1. Managerial impression management strategies in corporate narrative documents (Merkl-Davies & Brennan, 2007).

Readability/Reading Ease Manipulation

Readability can be understood as the ease at which one can comprehend a message in relation to the style of writing (Klare, 1963). An assessment of the writing style can be an appropriate and reliable measure of the difficulty associated with reading comprehension (Courtis, 2004). In order for the information in a message to be transferred effectively, it is dependent on the ability of the reader

to be able to understand the information provided. If the information is presented in a way which is too complex and beyond the comprehension capacity of the targeted audience, it can lead to the outcome of misunderstanding, impairment of reporting quality, or curtailment of the decision-making abilities of the intended audience (Courtis, 1995). One way of deliberately making the level of reading comprehension difficult is through reading ease manipulation. Reading ease manipulation is an attempt made to obscure a negative event by making the readability of documents difficult to comprehend (Merkl-Davies, 2007). When a firm is faced with poor performance or an undesirable situation, managers may attempt to conceal this information (e.g., by reducing its overall clarity) as a way of sidetracking the reader from the seriousness of the event.

As a way of measuring reading difficulty, a common method adopted is to use a readability formula. On the one hand, some researchers have adopted a single measure, for example, Flesch (Baker & Kare, 1992; Jones, 1988); Fog (Parker, 1982) and Cloze (Adelberg, 1979), while others have adopted a multi-method approach, including Flesch with some other combination of measures, such as, Fog, Dale-Chall, Lix, Fry, Cloze, word length (Courtis, 1986; Li, 2008; Smith & Taffler, 1992). Within the area of reading ease manipulation research, the dominant focus has been on annual reports, particularly the narrative of the chairman's statement (Courtis, 1995; 2004; Jones 1988; Li 2008; O'Donovan 2002). A common finding is that annual reports are difficult/very difficult to read (Courtis, 1995; 2004; Linsley & Lawrence, 2007). Often the approach followed is to compare the readability of annual report narratives with firm financial performance (Merkl-Davies & Brennan, 2007). However, findings appear to be inconclusive, with some studies (Courtis, 1986; Jones, 1988; Linsley & Lawrence, 2007) not finding support for the association between the reading difficulty of annual reports, while others have (Bakar & Ameer, 2011; Baker & Kare, 1992; Smith & Taffler, 1992).

Research gaps

Notwithstanding the importance of readability research, there have been increased calls acknowledging the need to extend our understanding of managerial impression management behavior in different corporate reporting contexts (Merkl-Davies & Brennan, 2007), as well as the need to move beyond the exclusive focus on firm financial performance and use other measures in conjunction with firm related variables. Considering different contexts of enquiry may offer opportunities to enhance our understanding of impression management strategies from the preparer perspective, in the next section, we examine the impact of data breach severity on the readability of mandatory data breach notification letters.

Research question

In summary, the study seeks to address the following research question: what is the impact of data breach severity on the extent of readability of the corresponding data breach notification letter(s), and under what conditions will managers employ a specific reading ease manipulation strategy?

Hypotheses

A summary of the research model is illustrated in figure 2.

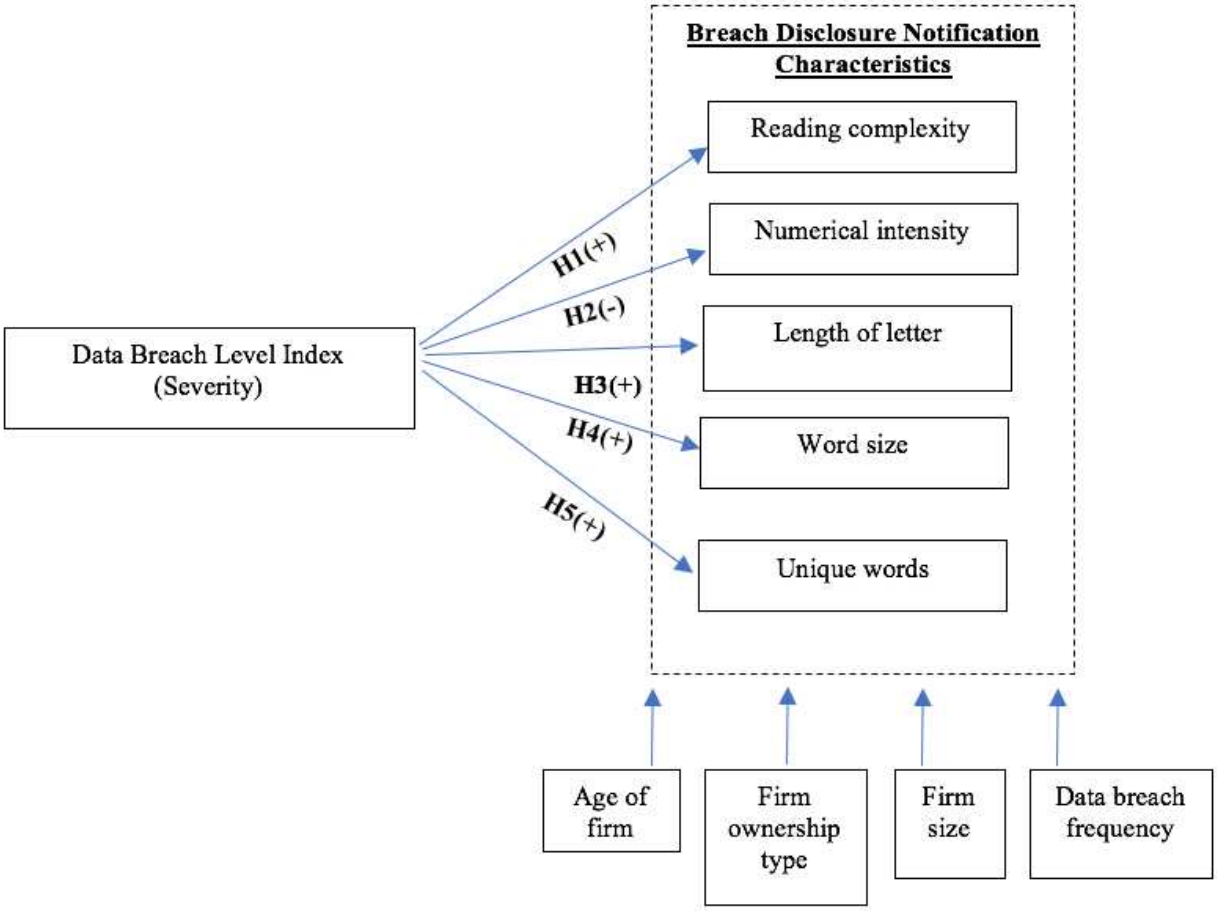

FIG. 2. Research model.

Reading Complexity

Reading complexity refers to how challenging a reader finds a passage to comprehend (Merkl-Davies, 2007). While reading complexity may be unintentional in nature, for instance, through poor writing practice (Courtis, 1995), from an impression management perspective, when firms are confronted with poor performance or a potentially damaging situation, the greater may be the propensity for business managers to obscure bad news by deliberately making the message more difficult to read (Courtis, 2004). By attempting to reduce the overall clarity of the message, the aim is to leave the reader puzzled and even confused in the hope that it puts the reader off or discourages them from investigating further (Merkl-Davies & Brennan, 2007). Translating this thinking to data breach notification, given that data breach messages are more unclear and complex than what is desired, one could argue that managers, when confronted with a higher data breach situation, would

want to write in a more complex way in an attempt to put readers off from effectively deciphering the message. It is thus proposed:

H1: The higher the data breach severity, the higher the degree of reading complexity within U.S. organizational data breach notifications.

Numerical Intensity

Numerical intensity refers to the extent to which numerical (quantitative) terms are present in the message conveyed (Henry, 2008). While the presence of numerical terms can be associated with exactness and enhanced credibility (Botosan, 1997; Mercer, 2004), from an impression management standpoint, particularly when communicating bad news, managers may prefer to use words over numbers. One possible reason for this is that making tangible (verifiable) numerical claims may draw closer reader scrutiny to the negative event (Craig, Mortensen, & Iyer, 2013); whereas words which are softer (strongly tied to emotions and values) in nature (Lipkus & Hollands, 1999), provide more opportunity for vagueness and generality (Demers & Vega, 2008). In the case of data breach communication, reduced reluctance to use numerical terms, when confronted with high data breach severity, may be treated as a mechanism to disguise the true intentions and underlying motives of managers by downplaying the use of specific and quantifiable claims. It can thus be hypothesized that:

H2: The higher the data breach severity, the lower the degree of numerical intensity within U.S. organizational data breach notifications.

Length of Letter

Overall length refers to the total number of words analyzed in a document and may be used as an impression management tactic (Cheung & Lau, 2016; Conway, O'Keefe, & Hrasky, 2015). One argument is that longer letters are used as a device to obscure negative news (Conway, O'Keefe, & Hrasky, 2015). As longer documents require more information-processing capability, this can act as a disincentive for recipients to read the total document (Loughran & McDonald, 2014) and be used to the advantage of business managers by attempting to hide adverse or bad news from stakeholders by burying the results in lengthier documents (Wallsten, Budescu, & Zwick, 1993). Given that much of the guidance on data breach communication advises that notifications should be concise and to the point, it may be the case that business managers, particularly when faced with higher data breach severity, would want to hide bad news in longer documents or having lengthier letters would discourage the recipient from reading all elements of the message. It is therefore postulated:

H3: The higher the data breach severity, the longer the letter within U.S. organizational data breach notifications.

Word Size

Word size can be understood as the number of characters per word in a given document (Pitler & Nenkova, 2008). Research (Carver, 1976; Coleman, 1971) has shown that there is a high correlation between reading difficulty and word length. In other words, when the word size becomes

larger, the more difficult the readability of the document becomes. Courtis (1995) makes the point that improvements in readability can be enhanced through a deliberate attempt by those responsible for crafting the communication by writing in a form which uses shorter words instead of longer ones. Applying word size to the study of data breach notification, particularly from an impression management perspective, one might expect to find larger word size to be associated with higher levels of data breach severity. The rationale would be that the use of larger word size compared to a shorter word size may lead to difficulties associated with reader interpretation, that is, larger words are often difficult for readers to evaluate, which can result in diverse interpretations by different readers, as well as inherent vagueness and ambiguous communication (Carver, 1976). We therefore postulate that:

H4: The higher the data breach severity, the higher the word size within U.S. organizational data breach notifications.

Unique Words

Unique words can be defined as the total word count with no duplicates counted. Research (Graesser, McNamara, Louwerse, & Cai, 2004; Lorge, 1949) has shown that unique words (e.g., uncommon words or technical terminology) can be associated with reading difficulty. More precisely, the higher the instances of unique words the higher the reading difficulty (Loughran & McDonald, 2014; Pitler & Nenkova, 2008). In comparison, writing which does not use, or uses less, unique words can be, in relation to reading comprehensibility, perceived as being more concise and simple (Das & Mukhopadhyay, 2006). Unique terms can act as a reading barrier in that messages which consist of many one-off vocabularies (which in many cases give rise to a convoluted sentence arrangement or prepositional phraseology) need time to be effectively deciphered and assimilated by the reader. In the context of data breaches, since uniqueness can imply a complex syntactical structure, managers, particularly when faced with higher data breach severity, may tactically use unique words (as opposed to more common words) as this makes the message more difficult to read. We therefore propose:

H5: The higher the data breach severity, the higher the unique words within U.S. organizational data breach notifications.

Data and Methodology

Data Description

Given that data breach reporting is now required in all 50 U.S. states and regulation is still awaiting or in the early stages of being implemented in different countries/regions across the world (e.g., Australia, Europe, Netherlands), it was decided to focus on the U.S. As state law requires private entities to notify residents that unencrypted personal information has been subject to a breach incident, in some states, a copy of the data breach letter, together with other relevant information (e.g., company name, date of data breach, date reported etc.) are publicly available online to read and download for educational purposes, usually via the Office of the Attorney General. Data collection was conducted during the Fall of 2016. Given the ease of availability of letters, the following Office of the Attorney General websites were selected for the investigation: California, Iowa, Maryland, New Hampshire and Vermont. To ensure consistency of the reporting period, letters received were selected from 2012 – 2015. The rationale for selecting this year range was that each of

the Office of the Attorney General websites had reported data breaches for each year of this period. All available letters, together with relevant information (e.g., company name, date of data breach, data breach reported date, state), were downloaded from across all websites.

In order for a letter to be deemed viable, it had to meet certain criteria. First, there had to be adequate information to allow for the calculation of the data breach severity score, as well as other relevant background information pertaining to the organization. If this was not available on the Attorney General website or present in the letter, a manual search was conducted to find this information. This involved searching the company website, media and other data breach reporting websites (e.g., www.privacyrights.org, www.databreaches.net). To establish validity of the data breach incident, the date of the data breach reported in the letter needed to be an exact match to the date stated in the source. Second, only letters which consisted of a signature from a member of the company holding a managerial position were considered for the investigation. One might raise the argument that data breach notification letters are crafted by professional writers and editors or, in many cases, the state provides a standardized template which businesses can adopt and use. While this might be the case, management values may manifest itself through analyzing communicative language, such as letters (Amernic, Craig, & Tourish, 2010), and given that the signature of the business manager features on the letter, it seems unlikely, given the reputational consequences, that managers would be unaware of the content of the letter.

From the initial letters downloaded, due to the choice of research design and other methodological issues (e.g., missing data, lack of signature), our sample consisted of 281 firms. As each firm may have multiple data breach incidents over the period examined, as well as several versions of each data breach letter, the final complete dataset consisted of 512 observations. Since additional information in the form of identity theft protection was frequently appended to letters and did not always represent the words of business managers, subsequently, it was decided that for each letter, any information below the signature would not be included in the analysis. All letters were in American English.

Description of Variables and Methodology

It is important to note that the focus is on, what we refer to as, successful data breaches i.e., those which have had an impact on the firm whereby consumer records were breached. The rationale for so doing is that, in many cases, firms do not have to disclose unsuccessful attempts and are reluctant to report this information to avoid public scrutiny. To measure data breach severity (independent variable), a Breach Level Index (BLI), developed by IT-Harvest and SafeNet, was used. Rather than using a single proxy to measure data breach severity, BLI acknowledges that severity is contingent on numerous factors and each factor consists of weighted values which determines the overall severity score. These include total number of records breached; the type of data that was exposed, the source of the data breach and how the data was used. BLI does not assign an upper limit, rather it is open ended; however, the largest data breach to date remains under 10. Like scales used for earthquakes and volcanoes, the index is logarithmic (base 10). Finally, the BLI score can be classified into one of five breach severity levels: minimal (1-2.9), moderate (3-4.9), critical (5-6.9), severe (7-8.9) and catastrophic (9-10). Table 1 provides a breakdown of data breach score and characterization.

Table 1. Data breach severity score and characterization (adapted from Stiennon 2013)

Category	Breach Level Index Score	Characterization
5	9 - 10	Breach with immense long-term impact on organization, customers and/or partners. Very large amount of highly sensitive information lost. Massive notification process. Potentially existential financial loss for breached organization in remediation and related costs.
4	7 - 8.9	A breach with significant exposure to business, legal and/or regulatory impact. Large amount of sensitive information lost. Significant notification process costs involved and public image impact.
3	5 - 6.9	A breach with likely short to mid-term exposure to business. Legal and/or regulatory impact. Usually moderate sensitive information involved. Some breach notification and financial loss.
2	3 - 4.9	A breach with low-term business impact. Usually involves loss of records of semi-sensitive information. Limited breach notification and financial exposure.
1	1 - 2.9	A breach with little/no material effect. Breach notification required, but little damage done.

In relation to the dependent variables (readability attributes), rather than using a single measure of readability, and in line with our hypotheses, we propose that our understanding of readability can be enhanced by using a multi-measure approach. In this paper, the following attributes, based on combining readability measures from various studies (Carver, 1976; Henry, 2008; Loughran & McDonald, 2014), — reading complexity, numerical intensity, length of letter, word size and unique words — are used to measure readability. DICTION 7, a text analysis program which assesses the language tone, was used. In measuring for reading complexity the variable complexity was chosen. Adopted from the work of Rudolph Flesch (1951) this variable is based on the premise that "convoluted phrasing makes a text's ideas abstract and its implications unclear" (Diction Manual, 2014). To measure numerical intensity the variable "numerical terms" was selected. The key assumption is that larger amounts of numerical terms present in a document lead to specificity, taking away from the universality of the claim made. Length of letter was measured by total number of words analyzed. Word size was measured by taking the average word size. The word count with no duplicates counted represented the measure for unique words.

Several control variables were introduced to account for the effects of other factors, including: firm ownership type (private or public); age of firm, data breach frequency and firm size. A definition of variables is shown in table 2.

Table 2. Definition of variables.

Variable	Description
Data breach severity	1-10 index (IT-Harvest and SafeNet BLI).
Complexity	A measure of convoluted phrasing based on the Flesch method (DICTION 7).
Numerical terms	Any sum, date, or product specifying the facts in a given case (DICTION 7).
Total words analyzed	Natural log of total number of words (DICTION 7).
Average word size	Number of characters per word in a given document (DICTION 7).
Unique words	Natural log of the word count with no duplicates counted (DICTION 7).
Firm size	A generated index 1-8, based on number of employees (index no = no of employees) 1(1-10); 2(11-50); 3(51-200); 4(201-500); 5(501-1000); 6(1001–5000); 7(5001-10,000); 8(10,001+).
Age of firm	Number of years since firm establishment.
Private firm	1 if the firm is private, 0 otherwise.
Data breach frequency	1 if more than one data breach was reported in the period examined, 0 if one data breach.

Log transformation is applied to the variables "total words analyzed" and "unique words" in order to reduce extreme values in the data associated with large variable magnitudes, which can lead to highly skewed distributions.

Content analysis was conducted to classify data breach severity attributes and other information. Using a sample of 10 letters, two of the authors independently coded data breach severity attributes and company information to confirm coding reliability[2]. The content analysis procedure involved using Excel software to help organize the data. For each data breach incident, the following procedure was followed. First, to calculate breach severity using the BLI, the following inputs were recorded; total number of records breached; the type of data that was exposed (unknown, nuisance, account access, financial access, identity theft or existential data); the source of the data breach (unknown, hacktivist, accidental loss, lost device, stolen device, malicious insider, malicious outsider or state sponsored), and how the data was used (unknown, undefined, no action was taken, action was taken, publicly disclosed/exposed or used for financial gain). Second, other important information relating to each data breach and organizational characteristics were also captured. This included company name, date of data breach; state (California, Iowa, Maryland, New Hampshire, Vermont); industry (education, financial, healthcare, other, retail, service, technology, travel or hospitality); firm size; age of firm; ownership type and data breach frequency (see table 2 for a description of the latter four variables).

In order to test for the effect of data breach severity on various readability attributes, we employed a multivariate regression analysis technique, which estimates a regression model with

[2] Both researchers had considerable experience in content analysis/coding. Percentage of agreement was used to establish inter-rater reliability. A percentage range of 90% to 95% is recommended to establish inter-rater reliability (Burns, 2014). The calculated score for our study was >95%. Any discrepancies were discussed between the researchers until agreement was reached and formed the basis for the establishment of coding rules for the remaining study.

more than one dependent variable, using five equations below. We are primarily interested in what impact the data breach severity variable has on the outcome variables, while data breach frequency, age of firm, firm size and private firm act as control variables.

$$\text{Total words analyzed}_i = \beta_0+\beta_1\text{Data breach severity}_i+\beta_2\text{Data breach frequency}_i+\beta_3\text{Age of firm}_i+\beta_4\text{Firm size}_i+\beta_5\text{Private firm}_i+e_i \quad (1)$$

$$\text{Average word size}_i = \beta_0+\beta_1\text{Data breach severity}_i+\beta_2\text{Data breach frequency}_i+\beta_3\text{Age of firm}_i+\beta_4\text{Firm size}_i+\beta_5\text{Private firm}_i+e_i \quad (2)$$

$$\text{Unique words}_i = \beta_0+\beta_1\text{Data breach severity}_i+\beta_2\text{Data breach frequency}_i+\beta_3\text{Age of firm}_i+\beta_4\text{Firm size}_i+\beta_5\text{Private firm}_i+e_i \quad (3)$$

$$\text{Numerical terms}_i = \beta_0+\beta_1\text{Data breach severity}_i+\beta_2\text{Data breach frequency}_i+\beta_3\text{Age of firm}_i+\beta_4\text{Firm size}_i+\beta_5\text{Private firm}_i+e_i \quad (4)$$

$$\text{Complexity}_i = \beta_0+\beta_1\text{Data breach severity}_i+\beta_2\text{Data breach frequency}_i+\beta_3\text{Age of firm}_i+\beta_4\text{Firm size}_i+\beta_5\text{Private firm}_i+e_i \quad (5)$$

Descriptive Statistics

The results show that the mean data breach severity is 5.197, with minimum and maximum values being 1.300 and 9.800, respectively. A histogram illustrating a further breakdown of the distribution of the number of data breaches for each level of severity is illustrated in figure 3.

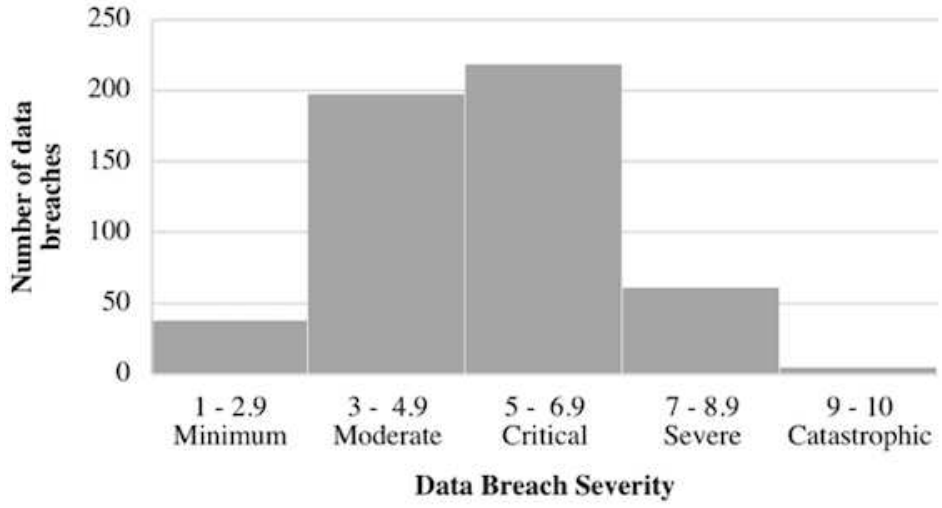

FIG. 3. Number of data breaches based on severity level

Reading complexity mean of the 521[3] data breach letters examined from 281 U.S. firms is 5.072. The mean of natural log of total words analyzed is 6.360, while the mean average word size is 5.075. Unique words show a mean natural log score of 5.654. Numerical terms display a mean of 27.742, however, we also note a high standard deviation, implying that the magnitude of this variable differs significantly from the mean across our sample. Considering firm characteristics, the average age of firm is 54.484 years, however, a high standard deviation once again points to the diversity of our sample. On average, firm size is recorded as 4.699, illustrating that firms employ between 200 and 1,000 workers. Thirteen percent of total data breaches reported by firms was not the first breach within the time period examined. Table 3 provides a summary of descriptive statistics.

Table 3. Descriptive statistics.

Variable	No. of obs	Mean	Standard deviation	Min	Max
Data breach severity	521	5.197	1.553	1.300	9.800
Total words analyzed (ln)	521	6.360	0.447	4.710	7.524
Average word size	521	5.075	0.206	4.330	5.570
Unique words (ln)	521	5.654	0.403	4.369	6.692
Numerical terms	521	27.742	14.669	0	111.500
Complexity	521	5.072	0.353	3.920	10.100
Firm size	512	4.699	2.357	1.000	8.000
Age of firm	512	54.484	50.529	3.000	262.000

Results

Correlation and Regression Analysis

Table 4 offers correlation coefficients for dependent and key independent variables. All independent variables display relatively low coefficients, signaling that multicollinearity will not affect the results produced by the multivariate models[4]. The preliminary results from the correlation table also show that there is a significant correlation between data breach severity score and all five dependent variables. Specifically, we note that there a positive correlation between data breach severity and total words analyzed, average word size, unique words and complexity, while there is a negative correlation between data breach severity and numerical terms.

[3] In total, 521 letters were examined. However, the final sample in the regression analysis was reduced to 512 due to several missing observations.

[4] We also re-ran regressions using a random effects model on panel data (generally, panel data approach is considered more suitable when there is a potential omitted variable bias) and determined that the produced results were consistent with our original findings. Fixed effects (panel) modelling was not possible due to a significant number of dummy variables in the dataset.

Table 4. Dependent and independent variables correlation coefficients (N=512).

	Total words analyzed	Average word size	Unique words	Numerical terms	Complexity	Data breach severity	Data breach frequency	Firm size	Private firm	Age of firm
Total words analyzed	1.000									
Average word size	-0.169*	1.000								
Unique words	0.988*	-0.161*	1.000							
Numerical terms	0.182*	-0.309*	0.181*	1.000						
Complexity	-0.130*	0.593*	-0.147*	-0.394*	1.000					
Data breach severity	0.175*	0.133*	0.160*	-0.164*	0.128*	1.000				
Data breach frequency	0.076	0.030	0.076	-0.019	-0.027	-0.033	1.000			
Firm size	-0.121*	0.115*	-0.126*	-0.056	0.095*	0.006	0.262*	1.000		
Private firm	0.054	-0.017	0.063	-0.002	-0.156*	0.034	0.007	-0.084	1.000	
Age of firm	-0.069	0.068	-0.082	-0.040	0.075	-0.105*	0.138*	0.321*	-0.202*	1.00

Note: * identifies correlation coefficients with p-values of 0.05 or lower.

Our regression analysis results are presented in table 5. Focusing on columns 1-5, it can be seen that data breach severity variable has a strong and statistically significant (at 1 percent level) effect on notification letter attributes[5]. Specifically, the regression analysis output identifies a positive impact of data breach severity on words analyzed, average word size, unique words and complexity, and a negative impact of data breach severity on numerical terms.

When the dependent variable is complexity, the coefficient for data breach severity variable displays a value of 0.029, providing support for hypothesis 1. In other words, higher data breach severity leads to higher complexity of the writing style. From our results, consistent with hypothesis 2, in relation to numerical terms, the data breach severity variable has a negative effect. Moreover, the variable coefficient has a value of -1.488, which is significantly larger than the effect of data

[5] STATA 2012 was used for the regression analysis. Before running the regressions, we have determined that multivariate analysis of variance results indicate that all five equations, when taken together are statistically significant and the independent variable simultaneously captures a statistically significant part of variance in the dependent variable.

breach severity on other dependent variables, highlighting the fact that business managers, when confronted with data breach severity, exhibit a communication preference of using words compared to numbers. In relation to hypothesis 3, the findings show that the data breach severity variable displays a coefficient of 0.047 when the dependent variable is total words analyzed (length of the letter). This indicates that the higher the data breach severity, the longer the data breach notifications tend to be. In line with hypothesis 4, the data breach severity coefficient shows a value of 0.016, indicating that higher data breach severity also results in higher average word size. However, we also note that the magnitude of this variable is relatively small and thus has a smaller impact on average word size, when compared to other readability measures. Finally, consistent with hypothesis 5, our findings indicate that the higher the data breach severity, the greater the use of unique words found in data breach letters (where data breach severity variable's coefficient is 0.039).

In addition, the results show that data breach frequency has a positive effect on total words analyzed and unique words (variable coefficients display substantial magnitudes of 0.153 and 0.139 respectively, both significant at 1 percent level). The effect of firm size on five dependent variables appears to be less clear. Larger firm letters tend to contain less words, as well as unique words, while having slightly longer words on average (although the impact of firm size on average word size is extremely small). Furthermore, our findings suggest that firm ownership has a strong negative impact on complexity (variable coefficient has a relatively large magnitude of 0.204, statistically significant at 1 percent level), signaling that privately-owned corporations tend to produce less complex letters. Lastly, we do not find any effect of firm age on all five data breach notification attributes.

We note from figure 3 that most data breach observations belong to a moderate (3 – 4.9) and critical (5 – 6.9) categories. In order to determine that these observations do not dictate our findings, we drop them from our dataset and rerun the regressions using the remaining observations (106 in total). However, generated results are still consistent with our previous findings, where the data breach severity variable has a positive impact on complexity, total words analyzed, unique words and average word size, and a negative impact on numerical terms.

While our model passes the required tests (e.g., the F-tests suggest that our chosen variables carry significant explanatory power), we note that the R-squared statistic is low, signaling that the changes in the predictor variables cannot always predict the change in the response variable. We therefore introduce a further robustness check, where we add three widely used additional control (dummy) variables into our regression analysis – relevant time period (year 2012, 2013, 2014 and 2015), US state (Maryland, Iowa, California, New Hampshire and Vermont) and industry the firm belongs to (retail, technology, health, education, finance, travel, services and other). Table 5, columns 6-10, display our main variables' coefficients and their significance levels once we control for the effects above. First, we can see that R-squared statistic has increased dramatically. Second, and more importantly, the results are consistent with our previously reported findings, highlighting a statistically significant effect of the data breach severity variable on notification letter attributes.

Table 5. The impact of data breach severity on readability attributes.

Variable	Total words analyzed (1)	Average word size (1)	Unique words (1)	Numerical terms (1)	Complexity (1)	Total words analyzed (2)	Average word size (2)	Unique words (2)	Numerical terms (2)	Complexity (2)
Data breach severity	0.047** (0.012)	0.016** (0.006)	0.039** (0.011)	-1.488** (0.417)	0.029** (0.010)	0.044** (0.013)	0.017** (0.006)	0.037** (0.012)	-1.541** (0.447)	0.033** (0.010)
Data breach frequency	0.153** (0.059)	-0.003 (0.028)	0.139** (0.053)	-0.223 (1.980)	-0.056 (0.047)	0.136* (0.060)	0.004 (0.027)	0.124* (0.054)	-0.864 (2.043)	-0.018 (0.048)
Firm size	-0.027** (0.008)	0.009* (0.004)	-0.025** (0.008)	-0.251 (0.295)	0.013 (0.007)	-0.020* (0.009)	0.010** (0.004)	-0.019* (0.008)	-0.328 (0.306)	0.013 (0.007)
Private firm	0.067 (0.078)	-0.001 (0.037)	0.071 (0.071)	-0.601 (2.621)	-0.204** (0.063)	0.095 (0.083)	0.065 (0.037)	0.092 (0.075)	-2.419 (2.845)	-0.081 (0.067)
Age of firm	-0.0001 (0.0004)	0.0002 (0.0002)	-0.0002 (0.0004)	-0.013 (0.014)	0.0003 (0.0003)	-0.0002 (0.0004)	0.00005 (0.0002)	-0.0004 (0.0004)	-0.012 (0.015)	0.0001 (0.003)
Time effect	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
US state effect	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Industry effect	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Constant	6.175** (0.110)	4.494** (0.051)	5.503** (0.099)	37.863** (3.682)	5.049** (0.088)	6.136** (0.206)	5.059** (0.092)	5.469** (0.186)	36.557** (7.028)	5.059** (0.165)
No. of obs	512	512	512	512	512	512	512	512	512	512
R squared	0.056	0.029	0.054	0.028	0.050	0.120	0.169	0.119	0.060	0.124
F-test	5.990**	3.022**	5.784**	2.963*	5.315**	3.522**	5.267**	3.500**	1.644*	3.662**

* and ** denote significance at the 5 and 1 percent level, respectively

Discussion and Conclusion

Referring back to our research question, the results of this study indicate that data breach severity has a positive impact on reading complexity, length of letter, word size and unique words, and a negative impact on numerical terms. Considered together, our results suggest that business managers, when confronted with higher data breach severity, may be engaging in impression management behavior by using a writing style which is more difficult to read (Merkl-Davies & Brennan, 2007). Reflecting on different theories of impression management introduced in the literature review section, we attempt to briefly offer some explanations for our results. Since data breach severity may result in lower company value, and the performance of managers is tied to financial incentives, in line with agency theory, managers may be partaking in reporting bias to avoid wealth reduction. While the focus of signaling theory is often on the managerial behavior of firms which perform well, who signal this superior performance by presenting and disclosing the information in a clearer manner, it may be the case, as illustrated in our study, that firms exhibiting lower data breach severity, are more transparent by making syntactical features of the message easier to read. Drawing on theories which concentrate more on environmental and social functioning (i.e., legitimacy theory, stakeholder theory, institutional theory), given that North American organizations, due to regulatory and legal compliance, are required to follow specific reporting procedures in the case of data breaches, unlike other types of IT security vulnerabilities, firms must be seen to conform to these societal and institutional norms of reporting. Nevertheless, managers may be merely complying (but make the message difficult to read) in an attempt to be seen to engage in corrective techniques which restores the relationship with society as a whole or specific stakeholder groups.

One might argue that much of the paper is written in a way which does not permit anything else but impression management as an explanation for the results, and alternative explanations may be at play. A simple alternative reason is that severe data breach incidents require more complex/longer letters to explain the situation. Given the nature of our results and the analytical lens used for the investigation, we cannot rule out this explanation. However, it seems unlikely that business managers would need to disclose information relating to data breach incidents in a more complex and lengthier way given that much of the data breach literature argues that the response should be discussed in clear and simple terms (Federal Trade Commission, 2016), and with many government sites providing recommended standardized templates, there isn't always a needed justification to use more words. Indeed, once we ran a regression analysis for only critical, severe, and catastrophic data breaches, the rationale being that higher data breach severity would be linked to longer letters and more complex vocabulary, we found that the data breach severity variable was insignificant and therefore does not impact readability.

In examining the overall logic of the impact of severity of the data breach on reading difficulty, one possible pathway, based on Merkl-Davies and Brennan's (2007) framework, might be summarized as follows: severity of successful data breach => a need to communicate the severity of a successful data breach to a firm's stakeholders => managerial self-serving motives => managerial impression management behavior (concealment) => type of information => type of manipulation => choice of a communication strategy (reading ease manipulation as an impression management strategy) => choice of reading difficulty by a manager across five major dimensions, including (1) reading complexity; (2) numerical intensity; (3) length of letter; (4) word size; and (5) unique words. However, it is important to stress that this is only a possible set of explanations for our current results, and further research is needed. A proposed extended model is illustrated in figure 4 (Appendix 1).

The study is not without limitations. The period examined was 2012-2015, and by extending the sample size and the period of years examined could reveal additional findings. Also, the research looks at a specific country (U.S.), and it cannot be claimed that the findings, and their implications,

can be generalizable and applicable to different countries. Another potential limitation of the study is the measures used for testing readability and data breach severity, other methods might uncover different results. For instance, public perception/reaction whereby an individual pays greater attention to how their data has been used, could also indicate data breach severity. Further studies could also examine if the type of company stakeholder (e.g., consumer, supplier, investor) affected by the data breach, and the number of affected stakeholders, influence the reading ease manipulation strategies adopted.

The paper raises a number of implications for society, practitioners (managers and regulators) and for the scholarly community. If mandatory data breach notification letters are being used as vehicles for impression management, rather than to provide consumers with transparent information, the quality and purpose of data breach notifications will be weakened. This runs the risk of consumers not being effectively informed of the facts pertaining to the data breach incident, as well as the inability to take appropriate and effective remedial measures. The challenge for managers is to shift the mentality from solely focusing on legal, financial and reputational consequences, which acts in their own self-serving interests, to encouraging measures which foster transparency and accountability. Regulators need to develop and promote educational programs to support firms, as well as other stakeholders, in the development of better ethical policies and clearer incident response mechanisms by training managers on what and how data breach notification letters should be comprised of. That failing, there is the need for stricter guidelines and/or better compliance with agency/regulatory requirements regarding simple and clear communication. In relation to the implications of the study for the scholarly community, this paper is a useful starting point for thinking about conceptual, empirical and methodological research between data breach severity and data breach notification narratives, as well as showing what, how, where and why the relationship deserves scholarly attention.

Finally, as more organizations become reliant on digital platforms, connected technologies, and mobile networks, the threat posed by data breaches will likely to continue. The findings discussed in this paper are important to get to know the motivations and values of business managers in their corporate communication response to data breach incidents. Managers should be aware of how their own biases can influence the readability of data breach communication and attempt to craft corporate communication in a way which is both genuine and beneficial to the true needs of organizational stakeholders. We hope that practitioners and researchers can utilize the findings of our study to think about managerial self-serving motives and impression management strategies associated with data breaches.

References

Abrahamson, E., & Park, C. (1994). Concealment of negative organizational outcomes: An agency theory perspective. *The Academy of Management Journal, 37*(5), 1302-1334.

Adelberg, A. (1979). Narrative disclosures contained in financial reports: Means of communication or manipulation. *Accounting and Business Research,* 10, 179-189.

Amernic, J., Craig, R., & Tourish, D. (2010). *Measuring and assessing tone at the top using annual report CEO letters.* Edinburgh, United Kingdom: The Institute of Chartered Accountants of Scotland.

Baginski, S., Hassell, J., & Hillison, W. (2000). Voluntary causal disclosures: Tendencies and capital market reaction. *Review of Quantitative Accounting and Finance, 15*(4), 371-389.

Baker III, H., & Kare, D. (1992). Relationship between annual report readability and corporate financial performance. *Management Research News, 15*(1), 1-4.

Bakar, A., & Ameer, A. (2011). Readability of corporate social responsibility communication in Malaysia. *Corporate Social Responsibility and Environmental Management, 18*(1), 50-60.

Botosan, C. (1997). Disclosure level and the cost of equity capital Christine A. Botosan. *The Accounting Review, 72*(3), 323-349.

Burns, M. (2014). How to establish interrater reliability. *Nursing, 44*(10), 56-58.

Carver, R. (1976). Word length, prose difficulty, and reading rate. *Journal of Literacy Research, 8*(2), 193-203.

Cheung, E., & Lau, J. (2016). Readability of notes to the financial statements and the adoption of IFRS. *Australian Accounting Review, 26*(2), 162-176.

Coleman, E. (1971). Developing a technology of written instruction: Some determiners of the complexity of prose. In E. Rothkopf, & P. Johnson, *Verbal learning research and the technology of written instruction* (pp. 155-204). New York, New York: Teachers College Press.

Connelly, B., Hoskisson, R., Tihanyi, L., & Certo, S. (2010). Ownership as a form of corporate governance. *Journal of Management Studies*, *47*(8), 1561-1589.

Conway, S., O'Keefe, P., & Hrasky, S. (2015). Legitimacy, accountability and impression management in NGOs: The Indian ocean tsunami. *Accounting, Auditing & Accountability Journal, 28*(7), 1075-1098.

Courtis, J. (1986). An investigation into annual report readability and corporate risk- return relationships. *Accounting and Business Research, 16*, 285-294.

Courtis, J. (1995). Readability of annual reports: Western versus Asian evidence. *Accounting, Auditing & Accountability Journal, 8*(2), 4-17.

Courtis, J. (2004). Corporate report obfuscation: Artefact or phenomenon? *The British Accounting Review, 36*(3), 291-312.

Craig, R., Mortensen, T., & Iyer, S. (2013). Exploring top management language for signals of possible deception: The words of Satyam's chair Ramalinga Raju. *Journal of Business Ethics, 113*(2), 333-347.

Das, S., & Mukhopadhyay, S. (2006). Readability modelling and comparison of one and two parametric fit: A case study in Bangla. *Journal of Quantitative Linguistics, 13*, 17-34.

Demers, E., & Vega, C. (2008). Soft information in earnings announcements: News or noise? *International Finance Discussion Papers 951.* Board of Governors of the Federal Reserve System.

Diction Manual. (2014). *Diction 7: The text-analysis program user's manual.* Retrieved June 1, 2017, from DICTION: http://www.dictionsoftware.com.

Elhai, J., & Hall, B. (2016). Anxiety about internet hacking: Results from a community sample. *Computers in Human Behavior, 54*, 180-185.

Ellison, N., Heino, R., & Gibbs, J. (2006). Managing impressions online: Self presentation processes in the online dating environment. *Journal of Computer-Mediated Communication*, *11*(2), 415–441.

Federal Trade Commission. (2016). *Data breach response: A guide for business.* Federal Trade Commission. Retrieved June 1, 2017, from business.ftc.gov.

Flesch, R. (1951). *The art of clear thinking.* New York: Harper.

Fullwood, C., Melrose, K., Morris, N., & Floyd, S. (2012). Sex, blogs, and baring your soul: Factors influencing UK blogging strategies. *Journal of the American Society for Information Science and Technology*, *64*(2), 345-355.

Goel, S., & Shawky, H. (2009). Estimating the market impact of security breach announcements on firm values. *Information and Management*, *46*(7), 404-410.

Graesser, A., McNamara, D., Louwerse, M., & Cai, Z. (2004). Coh-metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers, 36*(2), 193-202.

Harris, K. (2014). *California data breach report.* California: Attorney General California Department of Justice.

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

Jones, M. (1988). A Longitudinal Study of the Readability of the Chairman's Narratives in the Corporate Reports of a UK Company. *Accounting and Business Research, 18*(12), 297-305.

Klare, G. (1963). *The measurement of readability.* Ames, Iowa: The Iowa State University Press.

Krämer, N., & Winter, S. (2008). Impression management 2.0: The relationship of self-esteem, extraversion, self-efficacy, and self-presentation within social networking sites. *Journal of Media Psychology: Theories, Methods, and Applications, 20*(3), 106-116.

Li, F. (2008). Annual Report Readability, Current Earnings, and Earnings Persistence. *Journal of Accounting and Economics, 45*(2-3), 221-247.

Linsley, P. & Lawrence, M. (2007). Risk Reporting By The Largest UK Companies: Readability And Lack Of Obfuscation, *Accounting, Auditing & Accountability Journal, 20*(4), 620-627.

Lipkus, I., & Hollands, J. (1999). The visual communication of risk. *JNCI Monographs, 25*, 149-163.

Lorge, I. (1949). Readability formulae - An evaluation. *Elementary English, 26*(2), 86-95.

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

Mercer, M. (2004). How do investors assess the credibility of management disclosures? *Accounting Horizons, 18*(3), 185-196.

Merkl-Davies, D. (2007). *The obfuscation hypothesis re-examined: Analyzing impression management in corporate narrative report documents.* Bangor: Prifysgol Bangor University.

Merkl-Davies, D., & Brennan, N. (2007). Discretionary disclosure strategies in corporate narratives: Incremental information or impression management? *Journal of Accounting Literature, 26*, 116-196.

Min, J., & Kim, B. (2015). How are people enticed to disclose personal information despite privacy concerns in social network sites? The calculus between benefit and cost. *Journal of the Association for Information Science and Technology, 66*(4), 839-857.

Narisi, S. (2012). Companies fail to disclose data breaches to SEC. Retrieved June 1, 2017, from http://www.itmanagerdaily.com/companies-fail-to-disclose-data-breaches/.

NCSL, (2018). Security breach notification laws. Retrieved September 1, 2018, from http://www.ncsl.org/research/telecommunications-and-information-technology/security-breach-notification-laws.aspx.

O'Donovan, G. (2002). Environmental Disclosures in the Annual Report: Extending the Applicability and Predictive Power of Legitimacy Theory. *Accounting, Auditing & Accountability Journal, 15*(3), 344-371.

Parker, L. (1982). Corporate annual reporting: A mass communication perspective. *Accounting and Business Research,* 12, 279-286.

Pfeffer, J., & Salancik, G. (1978), *The External Control of Organizations: A Resource Dependence Perspective*. Harper & Row, New York.

Pitler, E., & Nenkova, A. (2008). Revisiting readability: A unified framework for predicting text quality. *EMNLP '08 Proceedings of the Conference on Empirical Methods in Natural Language Processing*, (pp. 186-195). Honolulu, Hawaii — October 25 - 27.

Ponemon Institute. (2014). *The afermath of a data breach consumer sentiment.* Ponemon Institute Research Report.

Ponemon Institute. (2018). *2018 Cost of a data breach study.* Benchmark research sponsored by IBM Independently conducted by Ponemon Institute LLC . Michigan: Ponemon Institute LLC.

Raban, D. (2009). Self-presentation and the value of information in Q&A websites. *Journal of the American Society for Information Science and Technology*, *60*(12), 2465-2473.

Rosenberg, J., & Egbert, N. (2011). Online Impression Management: Personality Traits and Concerns for Secondary Goals as Predictors of Self-Presentation Tactics on Facebook. *Journal of Computer-Mediated Communication, 17*(1), 1-18.

Sen, R., & Borle, S. (2015). Estimating the contextual risk of data breach: An empirical approach. *Journal of Management Information Systems, 32*(2), 314-341.

Smith, M., & Taffler, R. (1992). Readability and understandability: Different measures of the textual complexity of accounting narrative. *Accounting, Auditing & Accountability Journal, 5*(4), 84-98.

Statista. (2018). *Annual number of data breaches and exposed records in the United States from 2005 to 2018 (in millions).* Retrieved from Statista: https://www.statista.com/statistics/273550/data-breaches-recorded-in-the-united-states-by-number-of-breaches-and-records-exposed/.

Stiennon, R. (2013). *Breach level index.* Categorising Data Breach Severity with a Breach Level Index: Retrieved June 1, 2017, from http://breachlevelindex.com/pdf/Breach-Level-Index-WP.pdf.

Veltsos, J. (2012). An analysis of data breach notifications as negative news. *Business and Professional Communication Quarterly, 75*(2), 192-207.

Wallsten, T., Budescu, D., & Zwick, R. (1993). Comparing the calibration and coherence of numerical and verbal probability judgments. *Management Science, 39*(2), 176-190.

World Economic Forum. (2018). *Insight report: The global risks report (13th Edition).* Geneva: World Economic Forum.

Zaharopoulus, D., & Kwok, L. (2017). Law firms' organizational impressions management strategies on twitter. *Journal of Creative Communications*, *12*(1), 48–61.

Appendix 1

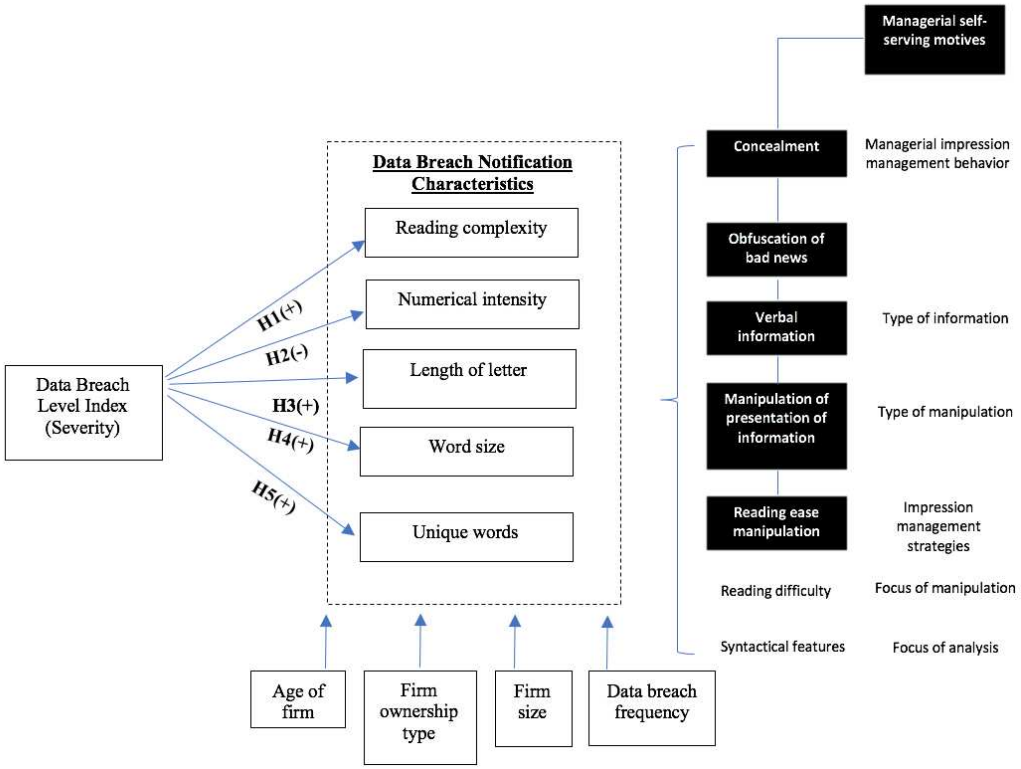

FIG. 4. Extended conceptual model.