Identification of fraudulent financial statements using linguistic credibility analysis

Sean L. Humpherys *, Kevin C. Moffitt, Mary B. Burns, Judee K. Burgoon, William F. Felix

University of Arizona, United States

ARTICLE INFO

ABSTRACT

The strategic use of deceptive language in managerial financial fraud is investigated with linguistic cues extracted from 202 publicly available financial disclosures. Those crafting fraudulent disclosures use more activation language, words, imagery, pleasantness, group references, and less lexical diversity than non-fraudulent ones. Writers of fraudulent disclosures may write more to appear credible while communicating less in actual content. A parsimonious model with Naïve Bayes and C4.5 achieved the highest classification accuracy. Results support the potential use of linguistic analyses by auditors to flag questionable financial disclosures and to assess fraud risk under Statement on Auditing Standards No. 99.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Despite the financial disasters of Enron, WorldCom, and Global Crossings, investors were shocked recently by the financial implosions of Lehman Brothers, AIG, Fannie Mae, and Freddie Mac. These cases underscore the need for investors and companies to protect their investments by detecting fraud in its earliest stages by distinguishing between truthful and misleading information. Investors look for credibility, transparency, and clarity in externally available corporate financial statements, such as the annually filed Form 10-K, as they investigate current and potential investments. This is especially true when financial markets are shaky.

The annual costs of corporate management fraud in the United States are estimated to be in the billions of dollars [57]. Fraud in general is "an act of deception carried out for the purpose of unfair, undeserved, and/or unlawful gain, esp. financial gain" [1]. Financial reporting fraud, also known as management fraud, is a type of fraud that adversely affects stakeholders through misleading financial reports [19]. Though the ability to identify fraudulent behavior is desirable, humans are only slightly better than chance at detecting deception [7], demonstrating the need for decision aids to help assess credibility. Thus, there is an imperative need for more reliable methods of identifying deception and fraud, especially in financial statements. New methods are needed to assist auditors and enforcement officers in maintaining trust and integrity in publicly owned corporations. Furthermore, investigations to detect deceit in financial statements can aid the overall investigation to refine general theories of deception.

One novel approach is to apply text-mining methods to the financial statements of companies. Ultimately, a decision aid based on these methods could help auditors assess the fraud risk of current and future clients. This study advances ongoing investigations into corporate fraud detection through a unique application of existing text-mining methods on the Management's Discussion and Analysis (MD&A) section of the Form10-K. The annually submitted Form 10-K is a required public company filing with the Securities and Exchange Commission (SEC) that "provides a comprehensive overview of the company's business and financial condition and includes audited financial statements" [52]. 10-Ks may contain fraud in the form of intentionally misstated numbers and/or misleading statements made by the authors. In the Form 10-K, a corporate annual report mandated by the Securities and Exchange Act of 1934, the MD&A section contains written explanations regarding the current status of the company, the industry, and forward looking statements for the company. Since the MD&A is intended to give investors a sense of management's perspective on the health and future outlook of a company, it contains a discussion of the company's financial condition, the results of operations, and an analysis of the quantitative and qualitative market risks facing the company. The MD&A, an unaudited section of the 10-K, is quasi-mandatory because much of the content is only suggested by the SEC and the content is largely uncontrolled. It is the most read section of the 10-K [50], but there is little research on the language used in the MD&A. Many scholars have called for additional research in this area [13].

The structure of this paper is as follows: we summarize current practices by auditors to detect deception in financial reports, review pertinent theories and methods for detecting deception and fraud, articulate our research questions, delineate our hypotheses, describe our methodology for detecting fraudulent financial statements, report the results, and discuss the implications of the findings.

---

* Corresponding author.
  E-mail address: shumpherys@cmi.arizona.edu (S.L. Humpherys).

2. Assessing credibility of financial statements

External auditors are tasked with planning and performing audits to obtain reasonable assurance about whether financial statements contain either inadvertent or intentional misstatements or omissions. As opposed to errors, intentional misstatements or omissions are part of Fraudulent Financial Reporting (FFR) meant to deceive users. Though problems in financial statements are introduced at various levels in organizations, FFR is most often committed by management. Under the Sarbanes-Oxley Act of 2002 (SOX), management, particularly the CEO and CFO, are not only responsible for creating the tone at the top for the corporate ethical culture, but are also accountable for discovering and preventing FFR in a publicly held entity.

Based on a well-planned and well-conducted audit, sufficient evidence is gathered for reasonable assurance that the risk of FFR is low but the risk is not eliminated completely. Due to concealment and/or collusion, fraud in financial statements/reports can be very difficult to detect. It is relatively rare for external auditors to find material misstatements or omissions [14,36,40]. Auditors must continually question and assess the audit evidence to maintain professional skepticism. To improve the audit processes associated with the detection of FFR, the American Institute of Certified Public Accountants' (AICPA) Auditing Standards Board (ASB) released Statement on Auditing Standard (SAS) No. 99 in 2002. Under SAS 99, auditors are required to take a more proactive approach to detecting FFR through improved and expanded audit procedures.

To identify the risk factors associated with each client, traditional audit techniques include enhanced analytical or statistical procedures, additional confirmation with external parties (e.g., customers) about unusual transactions or relationships, extra steps or observations to verify inventories, additional independent estimates to review management's estimates, and thorough review of financial data. Even with these additional procedures, auditors may not spot FFR. Therefore, specialized checklists or other procedures that augment the audit have been suggested by researchers and practitioners. For example, based on their study of SEC Accounting and Auditing Enforcement Releases (AAERs), Loebbecke et al. [40] devised a checklist of primary indicators or red flags for financial statement irregularities. These red flags are included in SAS 99. Schilit [47] described techniques for the hyper-skeptical auditor to spot major financial statement manipulation by management. Beneish [5] attempted to build a model based on extreme financial performance that identifies violators of Generally Accepted Accounting Principles (GAAP). His model successfully discriminated between fraudulent companies that experienced large positive accruals by manipulating their earnings and legitimate companies that are so-called "aggressive accruers." Logistic regression used to assess risk of FFR aid in classifying fraud vs. non-fraud engagements was helpful according to Bell and Carcello [4]. Kaminski et al. [31], focusing on a subset of Analytical Procedures (APs) used by auditors to augment typical audit procedures, found that financial ratios provide limited ability to detect FFR. However, Jones [29] identified other preliminary APs, such as market value of equity, that can help auditors assess fraud risk.

As new artificial intelligence and data mining technologies have become available, auditors have adopted some of these tools and techniques to help with fraud detection, primarily in examining the numerical data of financial statements. Gaganis et al. [25], Fanning and Cogger [20], Fanning et al. [21], Calderon and Cheh [12], and Lin et al. [39] examined the use of artificial and probabilistic neural networks for risk assessment of FFR. In 2004, Zhang and Zhou [58] reviewed various data mining techniques for financial and accounting applications such as credit card fraud detection. More recently, Kovalerchuk and Vityaev [35], Kotsiantis et al. [34], Kirkos et al. [32] applied various machine-learning techniques for data mining/classification of the financial data of FFRs. In other studies, Back et al. [3] and Kloptchenko et al. [33] mined both text and numerical data in a very limited set of financial statements for comparison, not fraud discovery, purposes. Though Minkin and Mosher [43] describe the use of message feature mining based on linguistic deception theories for processing e-texts, such as Enron's email, they do not suggest similar mining for FFR. The literature surveyed limited their investigations to numerical data, ignoring the text-based explanations that accompany the financial statements. However, the AAERs that accompany our collection of fraudulent FFRs identify evidence of deceptive communication, misdirection, and obfuscation in the text-based portions of the FFRs. This evidence suggests that the language in a FFR may be a fruitful area to investigate for fraud, especially if an automated tool can assist the auditor. In light of the lack of research on text and message feature mining of FFR, our research project offers a first step toward providing better audit risk assessment tools for auditors to detect FFRs. The current study complements past research that sought to discover numerical indicators of financial reporting fraud in financial statements [5,15,37,49] by evaluating linguistic cues of the MD&A section as indicators of financial reporting fraud. This study also investigates the usefulness of linguistic cues as a decision support model for credibility assessment.

3. Deception and fraud

Fraud is a form of deception. Deception is the act of transmitting information with the intent to foster false conclusions in the receiver [8]. Fraud "refers to an intentional act…to obtain an unjust advantage," but where there is no intent to deceive, error rather than fraud describes the act [27]. Fraud includes "a scheme designed to deceive" [56]. Management fraud is a specific type of deceptive scheme where stakeholders are adversely affected through misleading financial reports [19]. Since management fraud is a purposeful, strategic deception, behavioral deception theories and methods should help explain fraudulent behavior. This paper combines deception theory from Communication and Psychology literature with linguistic analysis techniques derived from the field of Computational Linguistics to understand the nature of the language used in fraudulent corporate SEC filings that are a traditional dataset in the field of Accounting. Prominent theories and methods for analyzing deceptive discourse include Content-Based Criteria Analysis (CBCA) [54], Scientific Content Analysis (SCAN) [17], Reality Monitoring (RM) [28], Management Obfuscation Hypothesis [6], Information Manipulation Theory (IMT) [42], Interpersonal Deception Theory (IDT) [8], Four Factor Theory [62], and Leakage Theory [18].

3.1. Content-Based Criteria Analysis

Content-Based Criteria Analysis (CBCA) is a method within Statement Validity Analysis, a technique developed to verify the veracity of a child's testimony in sex-crime cases. CBCA, however, has been used successfully in several different contexts. CBCA is based on the hypotheses that a statement based on fantasy will differ in quality and content from a statement based on actual experience. In CBCA, trained evaluators judge the presence or absence of 19 criteria. The presence of each criterion suggests that the statement was derived from an actual experience, and is therefore not deceptive. Deceptive statements should lack more criteria than truthful statements. Only some of the CBCA criteria are currently amenable to automatic analysis by computers including quantity of details, and words associated with feelings, time and space. CBCA hypothesizes that truthful messages will contain more unusual details, more superfluous details, more details overall, and more references to time, space, and feelings than deceptive messages because statements derived from actual memories of an experience should contain more contextual details than deceptive statements. It is uncertain, however, if these same cues will be of any significance in the context of managerial reports. For example, references to feelings may not appear at all in a managerial report.

Nevertheless, the MD&A section does allow for more diverse and less prescribed language than other parts of the 10-K so it will be informative to test for the existence and difference between the quantity of affect words in fraudulent and non-fraudulent 10-Ks. This same line of reasoning applies to the inclusion of any of the linguistic cues hereafter mentioned.

3.2. Scientific Content Analysis

Scientific Content Analysis (SCAN) is a unique version of CBCA that assumes that both deceivers and truthtellers are trying to convince the receiver of their truthfulness. Another important assumption of SCAN is that the sender carefully selects the details that enter into his/her account. This in turn suggests that each word is important in determining the veracity of a statement [17].

3.3. Reality Monitoring

Reality Monitoring (RM) is a method that attempts to distinguish between memories based on true experiences from internally generated falsehoods or imagination. In RM, memory is differentiated from imagination by its truthfulness. RM hypothesizes that statements based on true memories and statements based on falsehoods differ in the amount of perceptual details, the amount of contextual information, and the quantity of cognitive operations described in the statements. RM hypothesizes that truthful statements will provide more sound, visual, and tactile details than a false statement as well as more contextual references to time and location. False statements, on the other hand, should mention more cognitive operations than truthful statements. Cognitive operations are processes or acts of the mind such as thinking, admitting, understanding, and hoping that are used by deceivers to facilitate inventing false stories. Thus, an increase in the number of cognitive operations used in a statement should throw into doubt the veridicality of that statement.

3.4. Management Obfuscation Hypothesis

MD&A's are inherently difficult to read. However, according to the Management Obfuscation Hypothesis (MOH), MD&As that contain bad news should be even more difficult to read [6]. According to Bloomfield, MOH states that if management desires to delay market response to bad news then they will have an incentive to obfuscate or dissimulate the information. In other words, when companies perform poorly, management has an incentive to cover up this poor performance to delay stock price reaction by decreasing the readability [38] of their annual reports. This hypothesis is an extension of Bloomfield's incomplete revelation hypothesis that states that information that is more costly to extract and process will not be reflected immediately in market prices. Therefore, management can delay stock price reaction by obfuscating financial reports making them more costly to analyze. Managers of poorly performing companies wish to hide negative information by decreasing readability of financial reports [38]. Li used the SMOG reading index to operationalize and measure readability. The SMOG reading index takes into account sentence complexity (as measured by average sentence length) and word complexity (as measured by average word length) to determine readability. Longer sentences and longer words are a surrogate measure for complexity and should occur when fraud is present.

3.5. Information Manipulation Theory

Information Manipulation Theory (IMT) [42] bases its propositions on Grice's Conversational Implicature Theory, which gives four maxims for expectations from conversation. IMT states that deceivers covertly violate these maxims to dupe the receiver of the deception. The first maxim is the maxim of quantity, which relates to the amount of information that is shared in a message. It should neither be too little nor too much. Deceivers violate this maxim by withholding pertinent information while implying that they are sharing all of the information. Withholding information causes the receiver to be misled and come to an erroneous conclusion. The second maxim, the maxim of quality, relates to the veridicality of the information shared. People expect to receive information that is 100% true without compromise. Deceivers violate this maxim by inventing falsehoods and bold-faced lies that are meant to mislead the receiver. Thirdly, the maxim of relation is one of structure that dictates that responses should always relate to the preceding discourse. Deceivers violate this maxim by introducing extraneous information into a conversation making it difficult for the receiver to receive correct information about the topic they are investigating. Fourth is the maxim of manner that dictates that conversations should be brief, orderly, clear, and unambiguous. This maxim is violated by removing clarity from a conversation and replacing it with ambiguity.

3.6. Interpersonal Deception Theory

Although Interpersonal Deception Theory (IDT) is mainly concerned with deceptive interchanges dyadic and dialogic, the authors of IDT give many insights into how deceivers will behave in non-interactive, asynchronous settings in which documents like the MD&A occur [8]. Interpersonal Deception Theory merges principles from interpersonal communication and deception to deduce a series of 18 propositions that predict the behaviors of senders and receivers in an interactive context. Two important assumptions of IDT are that deception is goal-oriented and that deceivers want to minimize responsibility for their deceit if the deceit is discovered. This is especially important in the environment surrounding the content included in official financial statements for which the CEO and senior executives are personally held responsible by the SEC and Federal Government. Thus, to reduce the risk associated with the content of the documents that bear their signatures, managers would seek to minimize the number of stances or definitive statements made in those documents. This behavior would be more important and manageable in the MD&A section of the 10-K, a section that gives managers more flexibility and opportunity to say what they want to say. Another important assumption of IDT is that deceivers strategically manipulate information to attain their goals. This goal can be accomplished by managing a message's completeness, truthfulness, and/or relevance. IDT notes that deceivers and truthtellers alike try to manage their image, but companies engaged in fraud will more likely engage in image management by trying to falsely portray their company in a more positive light than is warranted. Non-fraudulent companies will be more likely to include negative information about their company if it is warranted. Finally, IDT notes that past research on non-interactive deception shows deceivers reduce specificity, use nonimmediate language, and use inclusive terms. These techniques will add ambiguity to statements and diffuse responsibility.

3.7. Four Factor Theory

The Four Factor Theory (FFT) describes the four processes or factors that influence deceivers' behaviors. The first factor is *Attempted Control*, which refers to deceivers controlling their behavior in an attempt to conceal their deception. The second factor is *Arousal*, which refers to the autonomic response of the nervous system in the deceiver at the time of the deception. The third factor is *The Affective Approach*, which refers to the emotions of guilt, anxiety, and duping delight that deceivers feel at the time of deception. These negative or unintended effects of deception may influence the deceivers to use nonimmediate language to disassociate themselves from the guilt induced by the deception. The final factor is *Cognitive Factors in Deception*, which refers to the increased cognitive load deceivers bear when inventing lies.

3.8. Leakage Theory

Leakage Theory (LT) describes unintentional behavioral cues that differentiate deceptive behavior from truthful behavior. These behavioral cues "leak" out because of a deceiver's inability to completely match behavior he or she would normally exhibit in a non-deceptive situation [18]. Examples of leakage by deceivers in face-to-face communication include an increase in shrugs, an increase in tension and fidgeting, a decrease in body and extremities movement, and a decrease in facial pleasantness [55]. However, these nonverbal cues are not present in written documents like 10-Ks, which make linguistic cues increasingly important for detecting deception. Also, financial documents are purposefully written with the quantity of and choice of words strategically employed as opposed to unintentionally exhibited.

In sum, these theories and methods provide a framework for understanding strategic and non-strategic deceptive behaviors of deceivers. As shown in Table 1, IDT, IMT, and MOH derive their propositions and hypotheses by focusing on deception as an intentional, strategic act. FFT and Leakage Theory focus on the unintentionally leaked cues during deceptive behavior. These approaches are not mutually independent since strategic behaviors are susceptible to leakage like any other behavior. Indeed, IDT, IMT, and MOH all posit the existence of leaked cues. CBCA, SCAN, and RM are tools for detecting specified leaked verbal cues. All of the theories posit that deceivers and truthtellers exhibit different verbal behavior. In addition, IDT, FFT, and Leakage Theory address kinesic and vocalic behavioral differences as well as differences in facial expressions.

This paper focuses on the linguistic cues to deception that may appear in fraudulent financial reports. Past research has already found that deceivers use different language than truthtellers [9]. For example, deceivers have been found to display elevated uncertainty, share fewer details, provide more spatio-temporal details, and use less diverse and less complex language than truthtellers [16,44,45]. Other research has found that deceivers use higher quantities of words, verbs, nouns, and group references and use more informal, non-immediate language than truthtellers [59,60]. In addition, research has shown that quantities of words and the use of specificity, affect, and activation terms differ depending on whether the deceiver begins a conversation by lying or by being truthful [10]. In their research, Zhou et al. [61] presented nine linguistic constructs useful for detecting deception, which are Affect, Complexity, Diversity, Expressivity, Nonimmediacy, Quantity, Specificity, Uncertainty, and Informality. Because Informality (e.g., misspelled words, typos) does not apply to a formal corporate report, we use only eight of these constructs in our current research to discriminate between truthtellers and deceivers.

The fact that deceivers have been found to use different linguistic cues than truthtellers provides an opportunity to potentially discriminate fraudulent from non-fraudulent financial statements. Because financial statements include management's explanation of the company's conditions and outlooks, we expect managers who

commit fraud in financial statements to display many of the same deceptive linguistic cues that have been observed in previous studies.

A common approach used to detect managerial fraud is to analyze numerical data to identify patterns of financial manipulation (e.g. [32,48]). We propose an alternative approach that analyzes the language contained in these same documents using natural language processing (NLP) techniques. NLP is a research area that focuses on using computing power to process natural language text [61]. NLP can be used to identify a variety of linguistic cues that then act as variables for statistical analyses or machine-learning algorithms. Both statistical and data mining techniques have been used to classify and predict deception using linguistic cues in non-financial related documents [24,26]. A careful analysis of textual features should reveal which linguistic cues discriminate documents containing deceit from truthful documents.

Along with contributing practical value by discerning fraud in financial statements, this study is designed to contribute to our understanding of the effects of major contextual variables in written deception. Findings by Zhou et al. [60] came in a laboratory setting where incentives to deceive were artificial and consequences of failure were minor or non-existent. Participants communicated via email to make a scenario-based decision. A confederate attempted to mislead the discussion using deception. Mann et al. [41] suggest that these low-stakes laboratory environments may not induce feelings of guilt nor elicit behavior found in realistic settings, thereby adversely affecting our ability to accurately judge credibility and diminishing the external validity of the results. Unlike deceptive laboratory email communications, managerial fraud can lead to serious consequences if the deception is discovered (e.g. reputation loss, financial loss, and/or incarceration). Also, in contrast to the Zhou et al. study [60], the communication modality of financial statements is essentially one-way and non-interactive. Compared to email or face-to-face communications, formal financial statements are prepared over a long period of time and can be rewritten and modified until they are deemed convincing. Thus, there is a need to discover whether linguistic analyses can accurately detect deception in real-world, high-stakes contexts where there is ample preparation time and little or no interactivity with the receivers.

4. Hypotheses

The annual filing of the 10-Ks is required by law to provide shareholders full disclosure of a company's profits, operations, and outlook. The MD&A section of the 10-K gives management an opportunity to give its perspective on the health of the company and its future outlook. If the outlook is bleak, or profits are lower than expected, the values of the company's shares will likely decrease. To avoid these losses, companies engaged in fraud might exclude negative news, include misleading positive statements, or create an optimistic outlook based on false premises to obfuscate the true state of the company. We would expect the language they employ to use more pleasant terms, more imagery, more affect, and more activation language in their MD&As than companies truthful about current conditions. In addition, managers will obfuscate their statements by increasing the complexity of their statements by writing longer sentences, using longer terms, and by increasing pausality. To deflect and diffuse responsibility for events or actions, we expect MD&As from fraudulent companies to include more nonimmediate language by referring to groups instead of individuals, mentioning others rather than themselves, and by communicating in the passive voice. Use of the passive voice (e.g., "mistakes were made by the company," and "performance was adversely affected") allows the filing company to disassociate itself from the message by either omitting the actor or making the actor the object of a statement. Words that indicate uncertainty, such as the modal verbs *would*, *should*, and *could*, lower the level to commitment to facts and assertions and should be used by deceiving managers to hedge against facts, assertions, and predictions.

Table 1
Summary of deception detection theories/methods.

Category	Theoretical foundation		Indicators of deception	
	Strategic deception	Leaked cues	Verbal cues	Nonverbal cues
Theories				
MOH	X		X	
IMT	X		X	
IDT	X		X	X
FFT		X	X	X
LT		X	X	X
Methods				
CBCA			X	
SCAN			X	
RM			X	

Due to the desire to be persuasive and appear credible without substantive details, those who craft deception in fraudulent 10-Ks will use more words overall while exhibiting less diversity and specificity of language than non-fraudulent 10-Ks. With more overall words, we also expect fraudulent MD&As to have more sentences, verbs, and modifiers than non-fraudulent MD&As. Specificity of language is measured by sensory words (words associated with the senses), spatial words (words that indicate either closeness or distance), and temporal words (words that either indicate present, future, or past relationships). Because the documents are written in a formal business language, we expect minimal use of self and group references. We expect an absence of typographical errors because all SEC filings should have been professionally proofread and errors removed. Formally stated,

Fraudulent MD&As display higher (a) quantity, (b) expressivity, (c) affect, (d) uncertainty, (e) nonimmediacy, and (f) complexity, and less (g) diversity and (h) specificity of language than non-fraudulent MD&As.

5. Methodology

To test for linguistic differences between fraudulent and non-fraudulent MD&As, we used the following subset of Zhou et al.'s constructs and variable definitions with modifications to the affect and spatial constructs.

Affect   *Activation Ratio*: number of activation words divided by the total number of words; *Affect Ratio*: Total number of affect words divided by the total number of words; *Imagery*: Number of imagery words divided by the total number of words; *Pleasantness Ratio*: number of pleasantness words.

Complexity   *Average Sentence Length:* Number of words divided by total number of sentences; *Average Word Length:* Number of syllables divided by total number of words; *Pausality:* Number of punctuation marks divided by total number of sentences.

Diversity   *Content Word Diversity:* Percentage of unique content words (number of different content words divided by total number of content words); *Function Word Diversity:* Number of function words divided by total number of sentences; *Lexical Diversity*: Percentage of unique words or terms out of total words.

Expressivity   *Emotiveness*: Ratio of adjective and adverbs to nouns and verbs.

Nonimmediacy   *Group References:* First person plural pronoun count divided by total number of verbs; *Other References*: Count of all other singular or plural pronouns divided by total number of verbs; *Passive Verb Ratio*: Number of passive verbs divided by total number of verbs.

Quantity   *Modifier Quantity:* Total number of modifiers; *Sentence Quantity:* Total number of sentences; *Verb Quantity:* Total number of verbs; *Word Quantity:* Total number of words.

Specificity   *Sensory Ratio:* Number of words referencing five senses, divided by total number of words; *Spatial Close Ratio, Spatial Far Ratio, Temporal Immediate Ratio, and Temporal Non-immediate Ratio:* Number of words that reference temporal or spatial information divided by total number of words.

Uncertainty   *Modal Verb Ratio:* Number of modal verbs divided by the total number of verbs.

Two of Zhou et al.'s original affect variables were not included because their lexicons for positive affect and negative affect were not available and they were not found to be statistically significant. Instead, we added Adkin et al.'s variables of pleasantness, activation and imagery. Adkins et al. also expanded the Zhou et al.'s single variable temporal-and-spatial ratio into four variables presented

above [6]. This composite model will be referred to as the 24-variable model. Twenty of the variables are ratio based and four are raw quantity counts. These variables are extracted using a decision support system called Agent99 Analyzer, which employs various classification algorithms after cue extraction.

The enforcement actions the Securities and Exchange Commission takes against firms that violate financial reporting standards are documented in Accounting and Auditing Enforcement Releases (AAERs). AAERs provide information regarding enforcement actions concerning "civil lawsuits brought by the Commission in federal court and notices and orders concerning the institutions and/or settlement of administrative proceedings" [51]. To conduct this study, we sought a sample comprised of discovered fraud cases as well as similar cases in which fraud had not been detected. We obtained the discovered fraud cases from the SEC's AAERs that were issued between 1995 and 2004. In our sample, we included AAERs that dealt with FFR, and more specifically, problems with the 10-Ks. The original 10-Ks, not the restated versions, were used in our fraud group.

The fraudulent 10-Ks were identified by searching for AAERs that included the term '10-K'. Companies named in AAERs are assumed to be guilty of earnings manipulations [15]. After excluding 40 companies and their associated 10-Ks from the 141 initially identified (see Table 2), 101 company 10-Ks remained for analysis. Table 3 summarizes the primary types of fraud found in the 10-Ks, as classified by the SEC's AAER.

We chose 101 comparable non-fraudulent 10-Ks by selecting companies with Standard Industrial Classification (SIC) codes that exactly matched the companies that filed fraudulent 10-Ks. After searching SEC's EDGAR (the database for online corporate financial information) for the same SIC code, companies within the same industry were randomly selected as potential non-fraudulent matches. Each matching company's 10-K was also filed in the same year or in the previous/following year and had no amendments. Those in the non-fraudulent group were verified as having no AAERs attached to them, which suggests a history of compliance with SEC regulations. The purposes of these criteria were to minimize potential confounds because of differing economic conditions between the fraudulent and non-fraudulent companies or to eliminate differences across dissimilar industries. In some instances the MD&A section includes tables of numerical data. The tables were excluded from the analysis to focus exclusively on linguistic cues.

Another possible confound was company size. There appear to be more large companies in the fraud group. Larger companies may produce longer 10-Ks or craft them differently than smaller companies. Larger companies may have more resources to apply towards hiding their deception. To investigate possible confounds of company size and total assets as possible significant explanatory variables of fraud or non-fraud, we performed a hierarchical two-model design using bivariate logistic regression. Total asset values were derived manually from each 10-K. The initial model resulted in a Nagelkerke R square of 0.269, explaining 26.9% of the variance in the dependent variable (fraud or non-fraud) and −2Log likelihood of 234.4 (where a lower score is better). If company size mattered, the Hosmer and Lemeshow test would have been significant and the predictive accuracy would have increased. Adding total assets to the initial model did not add any significant value in explaining the variance of the dependent variable according to a Hosmer and

Table 2
Selection criteria for fraudulent 10-Ks.

Count of companies identified as fraudulent by searching through AAERs	141
Count disqualified because fraud did not involve 10-Ks	(20)
Count disqualified because 10-K was not available from the SEC	(10)
Count disqualified because 10-K did not contain management discussion section	(10)
Final count of qualifying 10-Ks used in the final sample	101

Table 3
Primary types of fraud.

Type of fraud	Count of companies
Overstatement of revenues	44
Combination of overstating revenue and understating expenses	25
Disclosure issue	10
Overstatement of inventory	6
Other income increasing effects	6
Understatement of provisions for loan-loss reserves	5
Other	5

Lemeshow test (Chi-square 6.8, df = 8, p = 0.553) and failed to significantly improve the Nagelkerke R square (0.300) or −2Log likelihood (228.5) for the inclusive model. In fact, the predictive ability of the model actually decreased slightly when the total assets variable was added. Therefore, we concluded that even though large companies in the fraud group often write longer 10-Ks, company size does not negatively affect the creation of predictive models. To address the differential statement length, 20 of the linguistic variables were converted to ratios and therefore should describe deceptive behavior, not just verboseness. The remaining four variables are raw quantity variables.

6. Results

The 202 10-Ks were first submitted to Agent99 Analyzer, a part-of-speech tagger and text analysis tool [60] to automatically extract the 24 cues. The analysis was conducted as follows. Independent sample t-tests were performed. A parsimonious model was then constructed and multiple analyses of variance conducted to investigate difference between groups. Finally, classification techniques were evaluated. Both statistical and machine-learning techniques have been used to classify and predict deception based on text-based linguistic cues [24,26]. Past research has used machine-learning algorithms to identify deception in police person-of-interest statements [23], email [59], and chat-room logs [11]. We trained various machine-learning classification algorithms on the cues and tested their classification accuracy using a 10-fold cross validation as a bootstrap technique to increase validity of the results.

6.1. Testing the 24-variable model

As a first pass to identify possible discriminators of deception, each of the 24 cues were analyzed with one-tailed independent sample t-tests with alpha set to .05 for the Management's Discussion & Analysis, sections 7 and 7A of the 10-K. All observations were independent from each other in as much as each 10-K was written by a unique company and only one 10-K from each company was included in the sample. Mean values and significant differences are reported in Table 4 along with the hypothesized and obtained directions.

Truthful MD&A sections of the 10-Ks displayed greater content word diversity ($t_{(191.7)} = 4.643$, $p < 0.0005$), and lexical diversity ($t_{(200)} = 4.834$, $p < 0.0005$). Fraudulent MD&A sections displayed greater activation ($t_{(192.6)} = -2.785$, $p = 0.003$), imagery ($t_{(200)} = 2.837$, $p = 0.003$), pleasantness ($t_{(200)} = -2.738$, $p = 0.004$), average word length ($t_{(200)} = -4.212$, $p < 0.0005$), pausality ($t_{(200)} = -2.194$, $p = 0.015$), group references ($t_{(200)} = -2.064$, $p = 0.020$), modifier quantity ($t_{(200)} = 4.574$, $p < 0.0005$), sentence quantity ($t_{(200)} = -4.578$, $p < 0.0005$), verb quantity ($t_{(200)} = -4.482$, $p < 0.0005$), and word quantity ($t_{(200)} = -4.369$, $p < 0.0005$). The fractional degrees of freedom (DF) reported here for the t-tests indicate that the Levene's test for equality of variances was not able to be assumed and thus they were adjusted accordingly to preserve statistical validity.

6.2. Model reduction

To achieve greater parsimony and interpretability, data reduction techniques and the theoretical groupings of the variables were applied to the 24-variable model. Principal component factor analysis with Varimax rotation and reliability statistics guided the reduction. It resulted in a 10-variable model. Multivariate analyses of variance then tested the measures within each of the factors so as to examine them as a set and to reduce the probability of committing Type I errors. The reduced model is summarized in Table 5. Regarding the quantity construct, if word quantity is removed, the reliability measure increases to 0.919. However, this separation failed to produce any positive benefit to the classification models, so for simplicity and understandability, word quantity was retained in the quantity construct. The following variables did not share enough correlation with others to warrant merging into a higher construct: affect ratio, average word length, other reference, and pausality.

MANOVAs were conducted to investigate significant differences between fraudulent 10-Ks and non-fraudulent ones. The reduced Diversity construct was significant, Wilk's $\Lambda = 0.889$, $F_{(2,199)} = 12.4$, $p < 0.0005$, partial $\eta^2 = 0.11$, with deceivers demonstrating less diversity than truthtellers. Quantity was significant with deceivers writing more than truthtellers, Wilk's $\Lambda = 0.887$, $F_{(4,197)} = 6.25$, $p < 0.0005$, partial $\eta^2 = 0.11$. Active Language show near statistical significance and a noticeable effect size Wilk's $\Lambda = 0.940$, $F_{(6,195)} = 2.08$, $p = 0.057$, partial $\eta^2 = 0.06$. Active verb ratio is the weakest contributor to Active Language. Not surprisingly from the t-tests, the Sensorial construct (Wilk's $\Lambda = 0.988$, $F_{(2,199)} = 1.163$, $p = 0.315$, partial $\eta^2 = 0.01$) and Syntactic Complexity construct (Wilk's $\Lambda = 0.997$, $F_{(2,199)} = 0.34$, $p = 0.709$, partial $\eta^2 = 0.03$) fail to demonstrate statistically significant or exhibit noticeable effect sizes.

6.3. Classification algorithms

To identify classification accuracy of the linguistic features in distinguishing fraudulent from non-fraudulent 10-Ks, several statistical and machine-learning methods were utilized. The following machine-learning algorithms were selected to classify the 10-Ks because of their theoretically diverse foundation: C4.5 decision tree, Locally Weighted Learning (LWL), simple Naïve Bayes, and Support Vector Machine. Classification by logistic regression, a statistical technique, was also performed to provide comparison between statistical techniques and machine-learning algorithms.

Each algorithm builds a model based on a different set of theoretical premises. A decision tree algorithm results in a simple-to-understand tree graph with each leaf as a classification decision and each node as an attribute conjunction. We used the popular decision tree C4.5 algorithm [46], labeled J48 in Agent99 Analyzer. Locally weighted learning (LWL) algorithm was used with the following parameters: use all neighbors in width of weighting function, classifier is decision stump, no normalization, and linear weighting kernel [2,22]. A simple Naïve Bayes is a probabilistic classifier based on Bayes theorem. The support vector machine (SVM) normalizes all attributes and classifies by constructing hyperplanes in n-dimensional space. By using these theoretically diverse algorithms, the possibility of any one algorithm over-learning the data and not generalizing to a broader population was reduced.

Each algorithm was tested on the training set of fraudulent and truthful 10-Ks and tested using a 10-fold cross validation methodology (see Table 6), which is designed to yield greater generalizability and validity. A 10-fold cross validation is a bootstrapping technique that divides the data set into ten equal sets, uses nine sets to train the model, and uses the remaining set to test the model. This process is repeated ten times, with each set having one turn as the test set. The results from each of the ten tests are then averaged. Table 6 reports the results for each classification method.

Table 4
Linguistic cues analyzed by Agent99 Analyzer.

Construct and variables	Results	Non-fraud		Fraud	
		M	SD	M	SD
Affect					
Activation	F>N ***	1.647	0.022	1.655	0.018
Affect ratio	F>N	0.0041	0.003	0.0044	0.002
Imagery	F>N ***	1.476	0.044	1.492	0.032
Pleasantness	F>N ***	1.801	0.019	1.807	0.015
Complexity					
Average sentence length	F>N	20.30	3.000	20.57	2.318
Average word length	F>N ***	5.393	0.171	5.481	0.123
Pausality	F>N *	3.474	0.901	3.820	1.305
Diversity					
Content word diversity	N>F ***	0.361	0.102	0.300	0.083
Function word diversity	F>N	8.738	1.426	8.835	1.087
Lexical diversity	N>F ***	0.250	0.080	0.202	0.060
Expressivity					
Emotiveness	F>N	0.230	0.038	0.233	0.027
Nonimmediacy					
Group references ratio	F>N *	0.010	0.018	0.016	0.020
Other references ratio	F>N	0.00095	0.0009	0.00105	0.0007
Passive verb ratio	F>N	0.062	0.020	0.063	0.018
Quantity					
Modifier quantity	F>N ***	529	420	898	694
Sentence quantity	F>N ***	222	164	364	264
Verb quantity	F>N ***	613	485	1020	773
Word quantity	F>N ***	4612	3707	7603	5793
Specificity					
Sensory ratio	N>F	0.0594	0.010	0.0591	0.008
Spatial close ratio	N>F	0.011	0.004	0.010	0.004
Spatial far ratio	F>N	0.0456	0.009	0.0464	0.008
Temporal immediacy ratio	N>F	0.0026	0.0012	0.0024	0.001
Temporal nonimmediacy ratio	F>N	0.0018	0.0013	0.0018	0.001
Uncertainty					
Modal verb ratio	F>N	0.038	0.025	0.043	0.030

Note. F = fraudulent 10-Ks, N = non-fraudulent 10-Ks. The results column also indicates the direction hypothesized. One-tailed t-tests; *p-value<=0.05; ***significant at p-value<=0.005.

Bivariate logistic regression classified 128 instances correctly, 74 incorrectly; accuracy was 63.4% using a 10-fold cross validation (67.8% without cross validation); precision of fraud was 62.9%; precision of truthful was 63.9%; recall of fraud was 65.3% and of truthful 61.4%; F-measure for fraud was 64.1% and 62.6% for truthful; root mean squared error was 0.4335. The coefficients for the model are in Table 7. The −2Log likelihood was 242 and the Nagelkerke R^2 statistic indicated that the model accounted for 22.9% of the observed variance. To determine if greater parsimony could be attained, the 10-variable model was reduced to four variables based on the significance reported

Table 5
Reduced 10-variable model.

New variables	Averaged of values from old variables	Reliability
Active language	Activation, pleasantness, imagery, modal verb ratio, active verb ratio, and group reference	.675
Diversity	Lexical diversity and content word diversity	.979
Sensory terms	Sensory ratio and spatial far ratio	.900
Syntactic complexity	Function word diversity and average sentence length	.847
Temporal/emotiveness	Emotiveness, spatial close ratio, temporal nonimmediacy ratio, and temporal immediacy ratio	.039
Quantity	Verb, modifier, sentence, and word quantities	.511
Affect ratio	Affect ratio	NA
Average word length	Average word length	NA
Other references	Other references	NA
Pausality	Pausality	NA

Note. Cronbach's alpha reliability was used.

from MANOVA: diversity, active language, syntactic complexity, and sensory terms. A logistic regression produced 63.9% accuracy using a 10-fold cross validation, a −2Log likelihood of 252.9 and accounts for 16.8% of variances (Nagelkerke R^2). This finding provides evidence that, at least for the logistic regression, the four-variable model has the predictive accuracy equal to the 10-variable model and performs better than the 24-variable model.

Accuracy, recall, precision, and f-measures were calculated to evaluate the effectiveness of the machine-learning classifiers [30,53]. *Accuracy* is the total number of documents correctly classified divided by the total number of documents analyzed. *Recall of fraud* is the ratio of number of documents correctly classified as fraudulent to the total number of actual fraudulent documents. *Recall of truthful* is the ratio of the number of non-fraudulent documents correctly classified as non-fraudulent to the total number of actual non-fraudulent documents. *Precision of fraud* is the ratio of the number documents correctly classified as fraudulent to the total number of documents classified as fraudulent. *Precision of truthful* is the ratio of the number of documents correctly classified as non-fraudulent to the total number of documents classified as non-fraudulent. *F-measure* combines precision and recall into one metric using a weighted harmonic mean of each. As with all the metrics, a higher number denotes increased

Table 6
Classification accuracy of fraud/non-fraud.

Classification technique	24-variable model	10-variable model
Logistic regression	58.4%	63.4%
C4.5	64.9%	67.3%
LWL	66.3%	60.4%
Naïve Bayes	65.3%	67.3%
SVM	61.4%	65.8%

Note. 10-fold cross validation used for all tests.

Table 7
Coefficients of 10-variable model.

Variables	B
Diversity	− 2.4020
Active language	18.2710
Sensory terms	− 0.2140
Syntactic complexity	− 0.0290
Quantity	0.0002
Temporal/emotiveness	5.4540
Affect ratio	− 15.2970
Average word length	2.6440
Other references	− 122.2610
Pausality	0.2770
Constant	− 33.0900

Note. Modeled with logistic regression.

Fig. 1. Parsimonious decision tree for discriminating fraudulent 10-Ks.

effectiveness. Accuracy, precision, recall, and the F-measure are reported in Table 8 for each algorithm using the 10-variable model. Other specifics for each algorithm are reported as follows.

Feeding the 10-variable model into a C4.5 decision tree resulted in classifying 136 instances correctly, 66 incorrectly; accuracy was 67.3% using a 10-fold cross validation. Tuning the minimum number of instances per leaf to 37 maximized the accuracy. The decision tree is parsimonious requiring only two decision criteria (see Fig. 1). First, a quantity score is calculated and evaluated. Pursuant to the model-reduction phase of this research, the quantity score is calculated by summing counts of verbs, modifiers, sentences, and words then dividing by four (see Table 5). If the quantity score is less than 1060.75 the case is classified as non-fraudulent. This rule classifies 67 of the 202 cases. For the remaining cases, a second rule is applied. A score regarding active language is calculated by averaging the six individual ratios of activation, pleasantness, imagery, modal verb ratio, active verb ratio, and group reference (see Table 5). If the score for active language is less than 0.982 the case is classified as non-fraudulent, which affects 43 cases. If active language is greater than 0.982, the case is classified fraudulent. This rule affects the remaining 92 cases. Thus, one can conclude from the decision tree that fraudulent 10-Ks contain a higher quantity score and a higher active language ratio than non-fraudulent ones.

The false positive rate for non-fraudulent cases is 37.7% and the false negative rate for fraudulent cases is 30.7%. While we anticipate improved accuracies through further researcher, this balance between false positive and false negative rates indicates a lack of bias in the decision tree. This lack of bias is in contrast to human judgments, which exhibit a strong truth bias if the judge is a layman and a strong lie bias if the judge is a professional, such as an FBI agent, police officer, polygraph examiner, or other trained professional [7]. The other eight variables in the 10-variable model do not sufficiently reduce entropy to justify inclusion in the decision tree. When the other variables are forced into the decision tree, classification accuracy is reduced.

LWL classified 122 instances correctly, 80 incorrectly; accuracy was 60.4% using a 10-fold cross validation. It is the only algorithm to perform worse with the 10-variable model than the 24-variable model. Root mean squared error was 0.4985. Parameters are to use LWL with decision stump classifier, use all neighbors, nonnormalized, and the weighting kernel is calculated linearly.

Table 8
Accuracy, precision, recall, and F-measure using 10-variable model.

Classification technique	Overall accuracy	Precision		Recall		F-measure	
		Non-fraud	Fraud	Non-fraud	Fraud	Non-fraud	Fraud
C4.5	67.3%	66.7%	68.0%	66.7%	65.3%	68.0%	66.7%
LWL	60.4%	60.2%	60.6%	61.4%	59.4%	60.8%	60.0%
Jrip	67.3%	67.0%	67.7%	68.3%	66.3%	67.6%	67.0%
Naïve Bayes	67.3%	68.0%	66.7%	65.3%	69.3%	66.7%	68.0%
SVM	65.8%	67.8%	64.3%	60.4%	71.3%	63.9%	67.6%
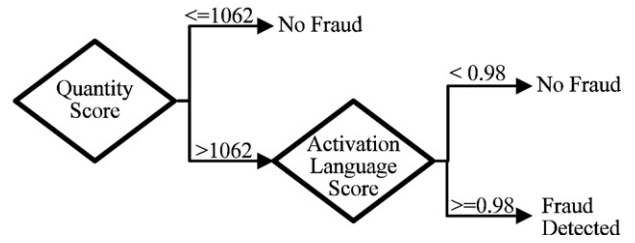
Note. 10-fold cross validation used for all tests.

To classify with JRip, we used the default Agent99 Analyzer settings of folds at three, minimum number of weight of the instances in a rule set at two, number of optimization runs at two, and pruning. The model resulted in the following: 136 instances were correctly classified, 66 incorrectly. Accuracy was 67.3% and the root mean squared error was 0.4717.

Simple Naïve Bayes was conducted with the default Agent99 Analyzer parameters of K2 search algorithm and Simple Estimator. It classified 136 instances correctly, 66 incorrectly; accuracy was 67.3% using a 10-fold cross validation. The root mean squared error was 0.4932.

SVM, using default parameters, classified 133 instances correctly, 69 incorrectly; accuracy was 65.8% using a 10-fold cross validation. The root mean squared error was 0.5845.

7. Discussion

Filing a fraudulent financial statement with the Securities and Exchange Commission is a serious criminal offense because the public relies on full disclosure of facts to make sound investment decisions. Large corporations involved in large-scale fraud, such as Enron, WorldCom, and Tyco, have seriously injured numerous individuals, costing many their retirement or livelihood. Full disclosure of the financial health of a company is critical to stable and efficient securities markets. Managerial fraud in financial statements is crafted with deceptive language because management seeks to hide or fabricate pertinent information about their company, which might negatively impact investment into the company or negatively affect management's performance-based compensation [56].

As a potential decision support tool, Agent99 Analyzer was used to extract pertinent linguistic features from the MD&A section of 101 fraudulent and 101 non-fraudulent 10-Ks. Two models of deception were investigated, a 24-variable model and a reduced model of 10-variables. Classification methods were used to evaluate the predictive accuracy of each model.

As hypothesized, fraudulent Management's Discussion and Analysis sections' language contained significantly more active language than non-fraudulent MD&As, as defined by activation word ratio, imagery, pleasantness, and modal verb ratio. Since fraud is often perpetrated in order to hide losses and meet Wall Street's expectations, managers may attempt to portray a false image of success by exaggerating positive news and minimizing or hiding negative news. Future research should create a measure of negative language to compare with the pleasantness construct from this study. When looking specifically at modal verbs, associated with communicating uncertainty, we found no difference in their use. This may be because all managers are encouraged by regulatory forces and generally accepted accounting practices to hedge their language.

We hypothesized that the fraudulent MD&As would contain more words but have lower lexical diversity. Managers perpetrating fraud are tasked to persuade readers of the veracity of their statements while distracting the reader from damaging information. One strategy is to generate quantities of irrelevant content [6]. This would have the effect of not only increasing the count of words, verbs, modifiers, and sentences in fraudulent MD&As, which was found to be a significant discriminator, but it would also dilute the diversity of language used

in MD&As. Our findings support this hypothesis. We found evidence that non-fraudulent MD&As had higher lexical word diversity and content word diversity than fraudulent MD&As, although there was no difference in function word diversity.

We predicted that fraudulent MD&As would be more complex than non-fraudulent MD&As. Our findings support the hypothesis that MD&As would have longer words and more pausality, but we found no difference in average sentence length. According to management obfuscation hypothesis, managers will try to make harmful information more difficult to extract in order to avoid or delay negative market responses. Deceitful managers can obfuscate the content of financial statements by creating longer documents and using more complex words.

Without decision aids, humans correctly classify lie–truth judgments only 54% of the time, slightly better than chance [7]. There is a scarcity of research that empirically quantifies an auditor's ability to detect financial statement fraud accurately. As reported in a recent KPMG Fraud Survey [36], only 12% of fraud was discovered by external auditors, while 65% was uncovered by the internal audit department. In 2002 the Accounting Standards Board issued Statement on Auditing Standards (SAS) No. 99, Consideration of Fraud in a Financial Statement Audit, to instruct auditors regarding their duty to detect fraud. While administering a survey to audit partners, Loebbecke, Eining, and Willingham [40] found that 112 out of 277 partners had no experience with respect to material irregularities. Since most external auditors have low experience in detecting fraud, finding decision aids to help auditors detect fraud is critical. Given that there are not any studies empirically quantifying an auditor's ability to detect financial statement fraud accurately, and given that humans can only successfully detect deception at an average of 54%, Agent99 Analyzer performed promisingly well with up to 67% accuracy in discriminating between fraudulent and non-fraudulent 10-Ks. Auditors can use Agent99 Analyzer, along with existing techniques, to screen and red flag potential fraudulent 10-Ks and thereby direct scarce auditing resources. In addition, future fraud detection research can use 67% as a base rate for quantifying improvements.

Guided by theoretical insight and exploratory factor analysis, the 24-variable model of deception was reduced to a 10-variable model. The reduced model had equal or better accuracy in classifying 10-Ks. A four-variable model was also created and successfully used in a classification test with logistic regression, which demonstrated that parsimonious models can perform as well as the larger models. The best performance came from the Naïve Bayes classifier and the C4.5 decision tree classifiers using the 10-variable model. Both achieved 67.3% accuracy. None of the models or classification technique suffered from a gross truth-bias or lie-bias, showing evidence of their balance between false positives and false negatives.

Our message feature mining study is the first to use qualitative, textual cues found in financial statements to discriminate between fraudulent and non-fraudulent cases. Hence, our study is foundational with respect to the text-based cues for deception detection in fraudulent and non-fraudulent financial statements. Therefore, this study provides benchmark findings against which future studies can be compared. Ideally, a computer aid such as Agent 99 Analyzer could pre-screen financial statements to enable auditors to focus efforts and resources more successfully.

8. Conclusion

The aim of this research is to expand our understanding of how deceivers use language differently than truthtellers, particularly in high-stakes, real-world environments such as financial markets. Natural language processing can help determine veracity by identifying textual cues that indicate the intent of the writer(s) in an organizational reporting context.

The modest success in classification results demonstrates that linguistic models of deception are potentially useful in discriminating

deception and managerial fraud in financial statements. Our findings provide critical knowledge about how deceivers craft fraudulent financial statements and expand the usefulness of deception models beyond a low-stakes, laboratory setting into a high-stakes, real-world environment where large fines and incarceration are the consequences of deception. These quantitative models and decision aids, like Agent99 Analyzer, could assist the SEC and financial auditors in detecting fraud and protecting the public's investments. Future research should further develop the deceptive models and expand investigation into other types of financial documents, such as the President's letter to shareholders, public announcements, glossy annual reports, and notes that accompany the financial statements.

References

[1] Anonymous, Fraud, [cited 2009 January 30, 2009]; Available from: http://en.wiktionary.org/wiki/fraud.
[2] C.G. Atkeson, A.W. Moore, S. Schaal, Locally weighted learning for control, Artificial Intelligence Review 11 (1–5) (1997).
[3] B. Back, J. Toivonen, H. Vanharanta, A. Visa, Comparing numerical data and text information from annual reports using self-organizing maps, International Journal of Accounting Information Systems 2 (4) (2001).
[4] T.B. Bell, J.V. Carcello, A decision aid for assessing the likelihood of fraudulent financial reporting, Auditing: A Journal of Practice & Theory 19 (1) (2000).
[5] M. Beneish, Detecting GAAP violation: implications for assessing earnings management among firms with extreme financial performance, Journal of Accounting and Public Policy 16 (3) (1997).
[6] R.J. Bloomfield, The "incomplete revelation hypothesis" and financial reporting, Accounting Horizons 16 (2002).
[7] C.F. Bond, B.M. DePaulo, Accuracy of deception judgments, Personality and Social Psychology Review 10 (3) (2006).
[8] D.B. Buller, J.K. Burgoon, Interpersonal Deception Theory, Communication Theory 6 (3) (1996).
[9] D.B. Buller, J.K. Burgoon, A. Buslig, J. Roiger, Testing Interpersonal Deception Theory: the language of interpersonal deception, Communication Theory 6 (3) (1996).
[10] J.K. Burgoon, T. Qin, The dynamic nature of deceptive verbal communications, Journal of Language and Social Psychology 25 (2006).
[11] J.K. Burgoon, F. Chen, D.P. Twitchell, Deception and its detection under synchronous and asynchronous computer–mediated communication, Group Decision and Negotiation 19 (4) (2010).
[12] T.G. Calderon, J.J. Cheh, A roadmap for future neural networks research in auditing and risk assessment, International Journal of Accounting Information Systems 3 (4) (2002).
[13] C. Cole, Management discussion and analysis: a review and implications for future research, Journal of Accounting Literature 24 (2005).
[14] C.P. Cullinan, S.G. Sutton, Defrauding the public interest: a critical examination of reengineered audit processes and the likelihood of detecting fraud, Critical Perspectives on Accounting 13 (2002).
[15] P.M. Dechow, R.G. Sloan, A.P. Sweeney, Causes and consequences of earnings manipulation: an analysis of firms subject to enforcement actions by the SEC, Contemporary Accounting Research 13 (1) (1996).
[16] B.M. DePaulo, J.J. Lindsay, B.E. Malone, L. Muhlenbruck, K. Charlton, H. Cooper, Cues to deception, Psychological Bulletin 129 (1) (2003).
[17] L.N. Driscoll, A validity assessment of written statements from suspects in criminal investigations using the scan technique, Police Studies: The International Review of Police Development 17 (4) (1994).
[18] P. Ekman, W.V. Friesen, Nonverbal leakage and clues to deception, Psychiatry 32 (1) (1969).
[19] R. Elliot, J. Willingham, Management fraud: detection and deterrence, Petrocelli, New York, NY, 1980.
[20] K. Fanning, K.O. Cogger, Neural network detection of management fraud using published financial data, International Journal of Intelligent Systems in Accounting, Finance & Management 7 (1) (1998).
[21] K. Fanning, K.O. Cogger, R. Srivastava, Detection of management fraud: a neural network approach, Proceedings of the 11th Conference on Artificial Intelligence for Applications, 1995.
[22] E. Frank, M. Hall, B. Pfahringer, Locally weighted naïve Bayes, Proceedings of the Conference on Uncertainty in Artificial Intelligence, Morgan Kaufman, 2003.
[23] C. Fuller, D.P. Biros, J.K. Burgoon, M. Adkins, D.P. Twitchell, An analysis of text-based deception detection tools, Proceedings of the 12th Americas Conference on Information Systems (AMCIS), Acapulco, Mexico, 2006.
[24] C.M. Fuller, D.P. Biros, D. Delen, Exploration of feature selection and advanced classification models for high-stakes deception detection, Proceedings of the 41st Hawaii International Conference on System Sciences (HICSS), Waikoloa, Big Island, HI, 2008.

[25] C. Gaganis, F. Pasiouras, M. Doumpos, Probabilistic neural networks for the identification of qualified audit opinions, Expert Systems with Applications 32 (1) (2007).

[26] J.T. Hancock, L.E. Curry, S. Goorha, M. Woodworth, On lying and being lied to: a linguistic analysis of deception in computer-mediated communication, Discourse Processes 45 (1) (2008).

[27] International Federation of Accountants, International Standard on Auditing 240: the auditor's responsibility to consider fraud and error in an audit of financial statements, , 2002.

[28] M. Johnson, C. Raye, Reality monitoring, Psychological Review 88 (1) (1981).

[29] K. Jones, Improving fraud risk assessments through analytical procedures, Marriott School (BYU) Accounting Symposium, 2005.

[30] M. Junker, R. Hoch, A. Dengel, On the evaluation of document analysis components by recall, precision, and accuracy, Fifth International Conference on Document Analysis and Recognition (ICDAR), 1999.

[31] K.A. Kaminski, T.S. Wetzel, L. Guan, Can financial ratios detect fraudulent financial reporting? Managerial Auditing Journal 19 (1) (2004).

[32] E. Kirkos, C. Spathis, Y. Manolopoulos, Data mining techniques for the detection of fraudulent financial statements, Expert Systems with Applications 32 (4) (2007).

[33] A. Kloptchenko, T. Eklund, J. Karlsson, B. Back, H. Vanharanta, A. Visa, Combining data and text mining techniques for analysing financial reports, Intelligent Systems in Accounting, Finance & Management 12 (1) (2004).

[34] S. Kotsiantis, E. Koumanakos, D. Tzelepis, V. Tampakas, Forecasting fraudulent financial statements using data mining, International Journal of Computational Intelligence 3 (2) (2006).

[35] B. Kovalerchuk, E. Vityaev, Data mining for financial applications, in: O. Maimon, L. Rokach (Eds.), Data Mining and Knowledge Discovery Handbook, Ch. 57, Springer US, 2005.

[36] KPMG, KPMG Fraud Survey 2003, KPMG, Montvale, NJ, 2003.

[37] T.A. Lee, R.W. Ingram, T.P. Howard, The difference between earnings and operating cash flow as an indicator of financial reporting fraud, Contemporary Accounting Research 16 (4) (1999).

[38] F. Li, Annual report readability, current earnings, and earnings persistence, Journal of Accounting & Economics 45 (2008).

[39] J.W. Lin, M.I. Hwang, J.D. Becker, A fuzzy neural network for assessing the risk of fraudulent financial reporting, Managerial Auditing Journal 18 (8) (2003).

[40] J.K. Loebbecke, M.M. Eining, J.J. Willingham, Auditors' experience with material irregularities — frequency, nature, and detectability, Auditing: A Journal of Practice & Theory 9 (1) (1989).

[41] S. Mann, A. Vrij, R. Bull, Suspects, lies, and videotape: an analysis of authentic high-stake liars, Law and Human Behavior 26 (3) (2002).

[42] S.A. McCornack, Information manipulation theory, Communication Monographs 59 (1) (1992).

[43] A. Minkin, D. Mosher, Linguistic deception theory: is the world of e-discovery ready? Deloitte Financial Advisory Services, 2008.

[44] M.L. Newman, J.W. Pennebaker, D.S. Berry, J.M. Richards, Lying words: predicting deception from linguistic styles, Personality and Social Psychology Bulletin 29 (5) (2003).

[45] T. Qin, J.K. Burgoon, J.P. Blair, J.F. Nunamaker Jr., Modality effects in deception detection and applications in automatic deception detection, Proceedings of the 38th Hawaii International Conference on Systems Sciences (HICSS), 2005.

[46] R. Quinlan, C4.5: programs for machine learning, Morgan Kaufmann Publishers, San Mateo, CA, 1993.

[47] H.M. Schilit, Financial shenanigans: how to detect accounting gimmicks and fraud in financial reports, 2nd ed.McGraw-Hill, New York, NY, 2002.

[48] C.T. Spathis, Detecting false financial statements using published data: some evidence from Greece, Managerial Auditing Journal 17 (4) (2002).

[49] S.L. Summers, J.T. Sweeney, Fraudulently misstated financial statements and insider trading: an empirical analysis, The Accounting Review 73 (1) (1998).

[50] L.R. Tavcar, Make the MD&A more readable, The CPA Journal 10 (January) (1998).

[51] U.S. Securities and Exchange Commission, Accounting and Auditing Enforcement Releases, [cited 2008 31 July 2008]; Available from: http://www.sec.gov/divisions/enforce/friactions.shtml.

[52] U.S. Securities and Exchange Commission, Form 10-K, 2006 Jan 27, 2006 [cited 2008 July 31, 2008]; Available from: http://www.sec.gov/answers/form10k.htm.

[53] C.J. van Rijsbergen, Information Retrieval, Butterworth, London, 1979.

[54] A. Vrij, Criteria-based content analysis: a qualitative review of the first 37 studies, Psychology, Public Policy, and Law 11 (1) (2005).

[55] A. Vrij, K. Edward, K.P. Roberts, R. Bull, Detecting deceit via analysis of verbal and nonverbal behavior, Journal of Nonverbal Behavior 24 (4) (2000).

[56] W.A. Wallace, Auditing, South-Western College Publishing, Cincinnati, OH, 1995.

[57] J.T. Wells, Occupational Fraud and Abuse, Obsidian Publishing, Austin, TX, 1997.

[58] D. Zhang, L. Zhou, Discovering golden nuggets: data mining in financial application, IEEE Transactions on Systems, Man, and Cybernetics — Part C: Applications and Reviews 34 (4) (2004).

[59] L. Zhou, J.K. Burgoon, D.P. Twitchell, A longitudinal analysis of language behavior of deception in e-mail, in: H. Chen, R. Moore, D. Zeng, J. Leavitt (Eds.), Intelligence and Security Informatics: First NSF/NIJ Symposium, ISI 2003, Tucson, AZ, USA, June 2–3, 2003 Proceedings, Springer, Berlin / Heidelberg, 2003.

[60] L. Zhou, J.K. Burgoon, J.F. Nunamaker Jr., D.P. Twitchell, Automating linguistics based cues for detecting deception in text based asynchronous computer mediated communication: an empirical investigation, Group Decision and Negotiation 13 (1) (2004).

[61] L. Zhou, J.K. Burgoon, D.P. Twitchell, T. Qin, J.F. Nunamaker Jr., A comparison of classification methods for predicting deception in computer-mediated communication, Journal of Management Information Systems 20 (4) (2004).

[62] M. Zuckerman, B.M. DePaulo, R. Rosenthal, Verbal and nonverbal communication of deception, in: L. Berkowitz (Ed.), Advances in experimental social psychology, Academic Press, New York, NY, 1981.

Sean Humpherys is a Ph.D. Candidate with the Center for the Management of Information Systems at Eller College of Management at the University of Arizona. His primary research interests include: Computer-assisted deception detection, HCI, machine-learning algorithms, data mining, and computational linguistics. He researches on projects funded by Department of Defense, Department of Homeland Security, National Science Foundation, Center for Identification Technology Research, National Center for Border Security and Immigration, and PwC.
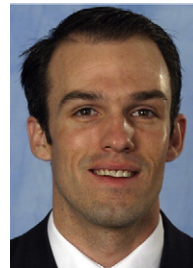
Kevin C. Moffitt is a Ph.D. Candidate with the Center for the Management of Information Systems at Eller College of Management at the University of Arizona. Kevin's research interests include: Automated deception detection applications, collaboration tools, accounting information systems, and communities of practice.

Mary B. Burns is a Ph.D. student with the Center for the Management of Information Systems at Eller College of Management at the University of Arizona. In addition to research interests in automated deception detection, Mary's research interests include: Business Process Reengineering (process/project workflow), IS Control/Project Control, IT Governance, and Data Security/Privacy.

Dr. Judee K. Burgoon is Professor of Communication, Professor of Family Studies and Human Development at the University of Arizona. She is the Director of Research for the Center for the Management of Information and Site Director for Center for Identification Technology Research at the University of Arizona. Professor Burgoon has authored or co-authored seven books and monographs and over 240 articles, chapters and reviews related to nonverbal and relational communication, interpersonal relationships, the impact of new communication technologies on human–human and human–computer interaction, research methods, and public opinion toward the media. Her research, which currently centers on deception, trust, interpersonal interaction, and new technologies, has been supported by extramural funding from the Department of Defense, the Department of Homeland Security, the U. S. Army Research Institute's Research and Advanced Concepts Office, the U. S. Army Research Office, the Office of Naval Research, the National Institutes of Mental Health, Gannett Foundation, Gannett Co., Inc., and Associated Press Managing Editors, among others.

Dr. William L. Felix, Jr. is the Eller Professor of Accounting at Eller College of Management at the University of Arizona. Dr. Felix's work is published in The Accounting Review, The Journal of Accounting Research, Journal of Accountancy, Auditing: A Journal of Practice and Theory, and other accounting journals. His primary expertise is in auditing and statistical methods in accounting and auditing.