Association for Information Systems
AIS Electronic Library (AISeL)

AMCIS 2009 Proceedings

Americas Conference on Information Systems (AMCIS)

2009

What Does That Mean? Investigating Obfuscation and Readability Cues as Indicators of Deception in Fraudulent Financial Reports

Kevin Moffitt
University of Arizona, kmoffitt@cmi.arizona.edu

Mary B. Burns
University of Arizona, mburns@cmi.arizona.edu

What Does That Mean? Investigating Obfuscation and Readability Cues as Indicators of Deception in Fraudulent Financial Reports

Kevin Moffitt
The University of Arizona
kmoffitt@cmi.arizona.edu

Mary B. Burns
The University of Arizona
mburns@cmi.arizona.edu

ABSTRACT

Building on theories of obfuscation and deception from accounting and communication literature, we examined 202 fraudulent and non-fraudulent 10-Ks by focusing on 25 linguistic cues. Our findings suggest that authors of fraudulent 10-Ks chose more complex words, signaling words of achievement and cause, and qualifying conjunctions. We found that truthful 10-Ks displayed more present tense verbs and were easier to read as indicated by the FRE readability measure. Those who construct 10-Ks may choose to deceive strategically by hiding bad news in more complicated content while trumpeting good news and achievements.

Keywords

Obfuscation, readability, fraudulent financial reporting, text mining, deception detection

INTRODUCTION

"FOOTNOTE 16. RELATED PARTY TRANSACTIONS:

In 2000 and 1999, Enron entered into transactions with limited partnerships (the Related Party) whose general partner's managing member is a senior officer of Enron. The limited partners of the Related Party are unrelated to Enron. Management believes that the terms of the transactions with the Related Party were reasonable compared to those which could have been negotiated with unrelated third parties…Subsequently, Enron sold a portion of its interest in the partnership through securitizations." (Enron Corporate Annual Report, 2000)

In the past year, the AIG, Lehman Brothers, Stanford International Bank, Freddie Mac, and Fannie Mae financial disasters mirrored those of the earlier financial crises of Enron, WorldCom, and Global Crossings. In unstable economies, investors seek reassurance before making or changing investments. Investors who want 'just the facts' about an organization turn directly to the back of the 'glossy' annual report to pore over sections such as the Management's Discussion and Analysis (MD&A) and the Footnotes. However, searching for clarity and credibility in financial statements can be extremely difficult. Though financial material can be difficult to read and comprehend in general (Courtis, 1998; Rutherford, 2003), obfuscation and convoluted language can be used both inadvertently and intentionally to mask the financial conditions of organizations. According to Billig (2008), nominalization and passivization of technical material may be an unconscious and ingrained attempt to exert control over information flow. In contrast, introducing unnecessary complexity in financial statements was a tactic of strategic misrepresentation used by the management of Enron as its annual reports, especially the footnotes (see above for infamous footnote 16), became increasingly longer and more convoluted over time.

To make annual reports and other documents filed with the SEC more approachable and understandable, the SEC developed "plain English" guidelines (U.S. Securities and Exchange Commission, 1998) Among the documents to which these guidelines should be applied is the Form 10-K which contains the Management's Discussion and Analysis, Items 7 and 7a. The MD&A contains text about the current and future financial state of a company as well as general information about its industry and is supported by quantitative analyses. Unlike other sections of the 10-K, the MD&A is less regulated by the SEC and is largely unaudited. Since the MD&As provide a rare glimpse into the current thinking of the top management of an organization, these Items have been prized, closely watched, and analyzed not only by investors but also by professional and academic researchers. While many aspects of the MD&A have been scrutinized by Accounting, Strategic Management, and Communication researchers, one focus has been to examine the relationship between readability and/or obfuscation and earnings persistence (Li, 2008), corporate profits (Courtis, 1986), sales (Jones, 1988), return on equity (Baker and Kare, 1992), and 'good' vs. 'poor' performance (Smith and Taffler, 1992a; Smith and Taffler, 1992b; Subramanian, Insley and Blackwell, 1993). Despite these studies, there has been a paucity of research of the association between the message features in texts of financial statements and fraud. To fill this void, we use linguistic credibility analysis tools to examine the correlation between cues of deception and both obfuscation and lack of readability in these texts. This paper describes theories of and prior research in both obfuscation/readability in financial statements and deception detection, poses our research questions and hypotheses, describes our methodology, examines our results, and presents a discussion of our findings, limitations, and future research directions.

OBFUSCATION AND READABILITY

In the Accounting literature, different hypotheses and theories (Incomplete Revelation Hypothesis (Bloomfield, 2002); the obfuscation hypothesis (Courtis, 1998); the management obfuscation hypothesis (Li, 2008); signaling and agency theories (Smith and Taffler, 1992a)) have been proposed to explain why managers choose less transparent reporting procedures as a way to boost stock price or improve bonuses. For example, according to the Incomplete Revelation Hypothesis (IRH), managers can make it harder for investors to discover bad news by burying that news in the footnotes while clearly showing, or even highlighting, good news in other parts of the 10-Ks. Consistent with obfuscation hypotheses, a manager purposely can make certain passages of financial reports less readable to obscure bad news while emphasizing good news. All of these hypotheses have roots in agency theory (Abrahamson and Amir, 1996; Abrahamson and Park, 1994; Aerts, 2005; Rutherford, 2003; Smith and Taffler, 2000) and in signaling theory (Rutherford, 2003; Smith and Taffler, 2000) which would predict that managers would downplay or conceal bad news and trumpet good news to increase the incentives of corporate officers and the overall value of the firm.

To study obfuscation hypotheses and to assess readability of text sections of annual reports, Courtis (1986, 1987, 1998, 2004), Smith and Taffler (1992b), Smith and Smith (1971), Soper et al. (1964), and Li (2008) used an assortment of psycholinguistic and sociolinguistic techniques such as CLOZE, the Flesch Reading Ease Index, the Fry, FOG, SMOG, Dale-Chall, Lix, and Rix measures. These measures, designed for general-purpose reading material and for readers of all ages, have shortcomings when applied to technical, specialized financial reports that are designed to be read by adults who are not novices. As Smith and Taffler (1992b) point out, traditional readability indexes may not be adequate for judging complexity of text. Our method of studying linguistic features of both obfuscation and readability will help to address that issue.

DECEPTION DETECTION

Intentional misstatements or omissions in financial statements are strategic schemes for deception. Researchers of deception detection (Buller and Burgoon, 1996; Ekman and Friesen, 1969) look for clues about ‘leakage’ that can discriminate between those who deceive and those who do not. Theories from Communication and Psychology, combined with linguistic analysis techniques from Computational Linguistics, inform our research. Interpersonal Deception Theory (IDT) (Buller and Burgoon, 1996) and Information Manipulation Theory (IMT) (McCornack, 1992) are the deception theories most relevant to this study. Interpersonal Deception Theory, originally posited for deceptive person-to-person exchanges, points to deception employed strategically, such as throwing a veil over a company’s financial condition by making its financial reports more obtuse. Primarily verbal and nonverbal cues for deception have been studied under the umbrella of this theory (Burgoon and Buller, 1994; DePaulo, Lindsay, Malone, Muhlenbruck, Charlton and Cooper, 2003; Ekman and Friesen, 1969), but related research has been conducted in both computer-mediated (Carlson, George, Burgoon, Adkins and White, 2004) and text-based communication (Fuller, Biros and Wilson, 2008; Newman, Pennebaker, Berry and Richards, 2003; Zhou, Twitchell, Qin, Burgoon and Nunamaker, 2003).

Information Manipulation Theory is based upon deceivers’ violation of the four maxims of Grice’s Conversational Implicature Theory. When violating the first maxim, deceivers hold back important information but indicate that they have disclosed fully. To breach the second maxim, deceivers misinform receivers who expect the unvarnished truth. To disregard maxim three in a conversation, deceivers respond in a complicated way to make it difficult to follow a topic thread. Deceivers who are unclear and unambiguous violate maxim four. Although these maxims were designed for a conversational mode of communication, the essence of these also applies to financial statements. Deceivers in text can employ a variety of techniques to omit key information, mislead, obfuscate, and/or provide roundabout explanations to covertly dupe the readers.

The theories described above differentiate between communications from deceivers and truth tellers. In part due to natural ‘truth bias’, humans, including those with special training, have disappointing results, slightly better than chance at 54%, in detecting deception (Bond and DePaulo, 2006). To improve that rate, other methods of deception detection such as automated linguistic cue analysis should be investigated. Previous research of automated deception detection in text concentrated on a variety of linguistic cues, such as quantity of words, that can signal deception (Zhou, Burgoon, Nunamaker and Twitchell, 2004; Zhou, Burgoon and Twitchell, 2003). Building on this previous research, our paper examines linguistic cues related to obfuscation and readability as indicators of fraud in financial statements. Thus, our research questions for this study are:

1. Do MD&A sections of fraudulent 10-Ks have a higher level of obfuscation?

2. Are MD&A sections of fraudulent 10-Ks associated less readable?

HYPOTHESES

Publicly traded companies must file a 10-K with the SEC on an annual basis so that shareholders, analysts, and other interested parties can read detailed information about the company, its financials, ongoing lawsuits, the company by-laws, and other legal documents with management's insights provided in the MD&A. Thus, the purpose of the 10-K is to provide full, rather than veiled, disclosure about a company's current status and its future prospects. However, if a company has disappointing results, is entangled legally, or has other bad news, there is goal incongruence between management's incentives to increase the company's value and SEC's reporting objectives. In a fraudulent situation, managers may decide to deceive through obfuscation or by making the document, as well as its underlying accounting procedures, unduly complex for many readers. Also, to distract the reader from disappointing news, management may try to highlight the good news. In order to distract readers, obfuscate, and make documents less readable, we expect fraudulent MD&As to have higher quantity of words and sentences, Word Complexity (longer words on average, higher rate of long words, higher rate of three+ syllable words), and Sentence Complexity (more words per sentence, more qualifying conjunctions). Qualifying conjunctions (e.g., except for, because) can create distance between the writer and the reader and make a sentence more complex. Also, we expect more: auxiliary verbs to construct passive voice, third person pronouns to blame outsiders or to deflect responsibility, cause words used to blame outsiders or outside influences for the bad news, and achievement-oriented words to emphasize the good news in fraudulent financial statements. Additionally, writers of fraudulent MD&As may focus more on the future as opposed to dwelling on explanations of the current situation and the past. They may also make the document less readable. We believe that truthful MD&As will use more first person plural and second person pronouns to adhere to clear writing guidelines. Formally, we hypothesize:

Fraudulent MD&As exhibit higher (a) Quantity (number of words, number of sentences), (b) Word Complexity (longer words, words with more syllables, more two+ syllable words), (c) Sentence complexity (more words per sentence, more qualifying conjunctions, (d) auxiliary verbs, (e) third person pronouns, (f) cause words, (g) achievement-oriented words, and (h) more future orientation; and, less (i) past and present orientation, (j) first person plural and second person pronouns,, and (k) readability than non-fraudulent MD&As.

METHODOLOGY

The enforcement actions the SEC 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" (U.S. Securities and Exchange Commission). AAERs often refer to required company filings as evidence of fraud and earnings manipulation.

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 1. Sample Selection Criteria for Fraudulent 10-Ks

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 (Dechow, Sloan, and Hutton, 1996). After excluding 40 companies and their associated 10-Ks from the 141 initially identified (see Table 1), 101 company 10-Ks were left for analysis.

101 comparable non-fraudulent 10-Ks were chosen by selecting companies with Standard Industrial Classification (SIC) codes that exactly matched the companies that filed fraudulent 10-Ks. Each matching company's 10-K was also filed in the same year or in the previous/following year and had no amendments. The purposes of these criteria are to minimize potential confounds because of differing economic conditions or differences between non-comparable industries. The non-fraudulent companies have no AAERs attached to them, which suggests a history of compliance to SEC regulations.

Software Tools

Linguistic Inquiry and Word Count (LIWC) is a computer program that counts the occurrences of words in different categories from an electronic document. For example, LIWC can count the number of words that indicate happy, sad, or angry feelings. LIWC has been used to study people's use of language in a variety of situations (Pennebaker and Graybeal, 2001). This study uses the words from the LIWC lexicon to measure many of the variables of interest.

LCExtractor (Linguistic Cue Extractor) is a tool written in Python 2.5 for automatically extracting linguistic cues from text, computing readability scores, and exporting the data to a CSV file. It has several functions that can easily be accessed by a novice Python user. LCExtractor, among other functions, extracts n-grams of any size, provides raw counts and ratios for custom-built lexicons, and counts syllables. Data are exported to a standard CSV file that can be imported into any spreadsheet and statistical software package. The readability indices that we used have each been studied in past research to some extent and include the Automated Readability Index (ARI), the Flesch Reading Ease formula (FRE), the Linsear-Write readability formula (LWRF), the Gunning-Fog Index (FOG), the Simple Measure of Gobbledygook index (SMOG), the Dale-Chall readability formula (DALE), the Lix readability formula (LIX) and the Rix readability formula (RIX).

Researchers have relied on different tools for automating the extraction of linguistic data from text. Recently, Fuller et al. (2008) used Agent99 and LIWC to automatically extract linguistic data that was analyzed with SPSS. Many researchers develop custom scripts in PERL or Python to extract linguistic cues (Li, 2008). Zhou et al. (2004) used a natural language processing tool called iSkim for identifying linguistic-based cues in conjunction with CueCal to derive a value for each linguistic cue.

RESULTS

We processed the text of the 202 MD&As with LIWC and LCExtractor. In a separate analysis, we found that LCExtractor identified sentence boundaries more accurately than LIWC. Thus, while average sentence length was calculated by LIWC and LCExtractor, the LCExtractor data was used. After the MD&As were processed, we conducted one-tailed independent sample t-tests to test for differences between the means of the variables for truthful and fraudulent MD&As. We conducted a correlation analysis for some word complexity and sentence complexity variables. We also analyzed the correlations between the readability indices.

The first step for identifying significant linguistic cues was to run a two-tailed independent sample t-test for each linguistic cue with a significance level of $p < .05$. We also identified near-significant constructs at a level of $p < .10$. We considered each observation from the MD&A to be an independent observation since each MD&A was written by a unique company. Table 2 reports the original hypotheses for each cue with the obtained direction, tscores, approximate p-values, means and standard deviations for each cue.

On average, fraudulent MD&As had more words overall ($t_{(200)}$=-4.545, $p < .001$), and more sentences ($t_{(200)}$=-4.907, $p < .001$). They also had more complex words as defined by average word length ($t_{(200)}$=-4.181, $p < .001$), and average syllables per word ($t_{(200)}$=-3.238, $p < .001$). They also had a higher rate of three syllable words ($t_{(200)}$=-2.077, $p < .05$). Fraudulent MD&As also used cause words ($t_{(200)}$=-2.220, $p < .05$), achieve words ($t_{(200)}$=-1.773, $p < .10$), and conjunctions ($t_{(200)}$=-2.231, $p < .05$) at a higher rate than truthful MD&As. Contrary to our hypotheses, our findings suggest that fraudulent MD&As contained more personal pronouns overall ($t_{(200)}$=-2.072, $p < .05$), including first person plural ($t_{(200)}$=-2.027, $p < .05$), and second person pronouns ($t_{(200)}$=-2.735, $p < .01$). Truthful MD&As used more verbs in present tense ($t_{(200)}$=1.938, $p < .10$), and were easier to read according to the FRE readability measure ($t_{(200)}$=1.846, $p < .10$) at a near-significant level. However, truthful MD&As used more auxiliary verbs ($t_{(200)}$=2.688, $p < .01$) than fraudulent MD&As which went against our hypothesis.

	Associated Variables	Predicted	Actual	NonFraud Mean	NonFraud Std Dev	Fraud Mean	Fraud Std Dev
Quantity		D>T	D>T***	4802.366	3755.910	8028.455	6064.407
	Number of Sentences	D>T	D>T***	253.564	184.032	415.475	275.890
Word Complexity	Rate of Six Letter Words	D>T	D>T***	33.921	2.783	35.216	2.175
	Average Word Length	D>T	D>T***	5.347	.171	5.438	.138
	Average Syllables per Word	D>T	D>T***	1.726	.077	1.758	.064
	Rate of Three Syllable Words	D>T	D>T**	.237	.025	.243	.021
Sentence Complexity	Average Sentence Length	D>T	T>D	19.837	5.112	19.346	4.038
	Conjunctions	D>T	D>T**	5.554	.818	5.796	.717
Personal Pronouns	First Person Plural	T>D	D>T**	1.002	1.700	1.513	1.876
	Second Person	T>D	D>T***	.003	.012	.012	.028
	Third Person Plural	D>T	D>T	.094	.085	.110	.070
Auxiliary Verbs	Auxiliary Verbs	D>T	T>D***	4.245	1.087	3.855	.974
Cause/Achievement	Causation Words	D>T	D>T**	2.550	.665	2.759	.676
	Achievement Words	D>T	D>T*	2.961	.894	3.186	.908
Tense	Past Tense	T>D	T>D	1.911	.740	1.760	.590
	Present Tense	T>D	T>D*	2.286	.839	2.091	.570
	Future Tense	D>T	T>D	.733	.427	.671	.368
Readability Measures	ARI	D>T	D>T	13.672	2.713	13.857	2.249
	FRE	T>D	T>D*	40.710	9.253	38.484	7.827
	LWRF	D>T	T>D	14.544	4.058	14.327	3.328
	FOG	D>T	D>T	31.595	3.597	32.078	3.147
	SMOG	D>T	D>T	15.414	1.816	15.471	1.580
	DALE	D>T	T>D	4.658	.255	4.635	.202
	LIX	D>T	D>T	63.577	5.622	64.154	4.838
	RIX	D>T	D>T	7.209	1.975	7.247	1.687

* = p < .10, ** = p < .05, *** = p < .01

Table 2. Linguistic Cues Analyzed by LIWC and LCExtractor

A correlation analysis (see Table 3) revealed that the number or words and sentences are highly correlated. Average word length, average syllables per words, and the number of words with six or more letters are also highly correlated.

	Word Count	Number of Sentences	Average Syllables per Word	Average Word Length	Rate of Six Letter Words
Word Count	1.000				
Number of Sentences	**.952**	1.000			
Average Syllables per Word	.365	.307	1.000		
Average Word Length	.282	.253	**.759**	1.000	
Rate of Six Letter Words	.244	.236	**.786**	**.824**	1.000

Table 3. Correlation Analysis of Theoretically Similar Variables

Readability Indices

We used eight readability measures to test the readability of the MD&As. Only the Flesch Reading Ease index indicated any difference in readability between fraudulent and truthful MD&As (at $p < .10$), fraudulent MD&As being more difficult to read. The ARI, FRE, LWRF, FOG, SMOG, LIX and RIX all indicated that the MD&As were meant for advanced readers. The FOG index, which is supposedly set to a grade level scale, gave an average score of ~32, equivalent to grade 32, for both corpora. The Dale-Chanell index was an anomaly in that it showed that the MD&As should be readable at a 4th grade reading level. Table 4 compares the readability scores from the MD&As to 'typical' readability scores for 5th grade and college level reading material.

	Scale	5th grade*	College Level*	Non-Fraud Score	Fraud Score
ARI	U.S Grade Level	5	13	13.67	13.86
FRE	0-100	90	30	40.71	38.48
LWRF	U.S Grade Level	5	13	14.54	14.33
FOG	U.S Grade Level	5	13	31.60	32.08
SMOG	U.S Grade Level	5	13	15.41	15.47
DALE	0-15	5	10	4.66	4.63
LIX	0-75	25	55	63.58	64.15
RIX	0-10	1.3	7.2	7.21	7.25

Table 4. Comparison of MD&A Scores and Typical Readability Scores

A correlation analysis showed the high correlation between the readability measures which means that the readability scores consistently went up or down according the difficulty of the document. The negative correlation of FRE with the other measures reflects the fact that in FRE higher scores indicate a document is easy to read. The results are summarized in Table 5.

	ARI	FRE	LWRF	FOG	SMOG	DALE	LIX	RIX
ARI	1.000							
FRE	-.854	1.000						
LWRF	.970	-.800	1.000					
FOG	.857	-.926	.832	1.000				
SMOG	.962	-.887	.971	.935	1.000			
DALE	.956	-.738	.991	.760	.936	1.000		
LIX	.970	-.865	.921	.855	.931	.900	1.000	
RIX	.985	-.829	.980	.853	.968	.968	.960	1.000

Table 5. Correlation of Readability Scores

Fraudulent MD&As should be more difficult to read because of managers trying to misinform, mislead, and misdirect their reading audience. Only one readability measure, FRE, supported the hypothesis that fraudulent MD&As are less readable than non-fraudulent MD&As. The other seven readability measures showed no difference. This may be due to the variables that were used to calculate the readability scores and weight each measure placed on the variables. As can be seen in Table 6, FRE is the only readability measure that incorporates average syllables per word. Five of the indices include average sentence length in their formulas and four include the rate of three syllable words. ARI was the only measure to include average word length, a highly significant variable. RIX is the only measure that relies on one variable for its score.

	Avg Word Length	Avg Sent Length	Avg Syl per Word	Rate Long Words	Rate 3 Syl Words	Number of Words
ARI	X	X				
FRE		X	X			
LWRF					X	X
FOG		X			X	
SMOG					X	
DALE		X			X	
LIX		X		X		
RIX				X		

Table 6. Items used to calculate readability measures

DISCUSSION

Our results suggest that higher Word Complexity (longer words and words with more syllables) is a feature of fraudulent MD&As. One strategy for obfuscating information that would lead to more complex words is nominalization of verbs, adjectives, and processes. Nominalized words are the noun form of verbs and adjectives and typically end with a "ment," "ance," "ence," “ness,” or "ion". Examples include “assessment”, “durableness”, and “nominalization”. Nominalized words are generally longer and more complex than other nouns which gives a possible explanation as to why fraudulent MD&As contain more complex words. Also, nominalizations are used to reify processes and verbs into entities and agents. Fowler (1991) added that nominalizations make text hard to understand and can be used to cover up the truth. However, nominalizations are not always ideologically motivated. Nouns may be derived over time from verbs and becomes standard lexical items such as ‘reporting’ from ‘to report’ and ‘reference’ from ‘to refer’. However, these types of nominalizations should uniformly occur in MD&As irrespective of fraudulent status since they are used non-strategically.

We found some evidence to suggest that fraudulent documents have more complex sentences because they include more qualifying conjunctions. Contrary to the recommendations in the SEC’s plain English writing guidelines, the fraudulent MD&As used more conjunctions such as ‘except’, and ‘because’. We believe there are two possible explanations for this: one is to create a more complicated sentence structure in which the logic is harder to follow; the second is to displace or deflect responsibility.

We hypothesized that fraudulent documents would contain fewer second-person pronouns and fewer first person plural pronouns than non-fraudulent documents because using these pronouns should lead to clearer, more understandable language. Contrary to our hypothesis we found that fraudulent MD&As used these pronouns more often than non-fraudulent MD&As. However, the rate at which second-person pronouns were used was so low (about .01% or less) that the importance of the measure is called into question.

As predicted, fraudulent MD&As did include more achievement ($p < .10$) and causation words. This suggests signaling behavior in that good news is trumpeted and bad news is attributed to others’ actions or external forces. Truthful MD&As used more present tense language than fraudulent MD&As. We predicted this outcome because truth tellers should be more comfortable explaining current results and ongoing operations. Contrary to our hypothesis, truthful MD&As used more auxiliary verbs than non-fraudulent MD&As. One explanation for this finding is that the use of auxiliary verbs in the progressive tense, such as “the company is growing 5% annually.” This use of auxiliary verbs complements the use of present tense in truthful MD&As.

LIMITATIONS AND FUTURE RESEARCH

Multiple authorship, or an established style for corporate communications, may be alternative explanations for our findings (Courtis, 2004). Another explanation for obfuscation in financial statements could be that bad news is harder to communicate, making the statements more complex overall.

We recognize that other financial reporting manipulations can be accomplished in charts, graphs, and/or supplementary financial analyses, but these are beyond the scope of this project. Future research can explore other text portions of the 10-Ks as well as related financial filings. Additional linguistic cues for detecting obfuscation need to be identified. Future studies might combine highly correlated variables into one representative construct.

REFERENCES

1. Abrahamson, E. and Amir, E. (1996) The Information Content of the President's Letter to Shareholders, *Journal of Business, Finance, & Accounting*, 23, 8, 1157-1182.
2. Abrahamson, E. and Park, C. (1994) Concealment of Negative Organizational Outcomes - an Agency Theory Perspective, *Academy of Management Journal*, 37, 5, 1302-1334.
3. Aerts, W. (2005) Picking up the Pieces: Impression Management in the Retrospective Attributional Framing of Accounting Outcomes, *Accounting, Organizations and Society*, 30, 6, 493-517.
4. Baker, H.E., III and Kare, D.D. (1992) Relationship between Annual Report Readability and Corporate Financial Performance, *Management Research News*, 15, 1-4.
5. Billig, M. (2008) The Language of Critical Discourse Analysis: The Case of Nominalization, *Discourse Society*, 19, 6, 783-800.
6. Bloomfield, R.J. (2002) The "Incomplete Revelation Hypothesis" and Financial Reporting, *Accounting Horizons*, 16, 233-243.
7. Bond, C.F. and DePaulo, B.M. (2006) Accuracy of Deception Judgments, *Personality and Social Psychology Review*, 10, 3, 214-234.
8. Buller, D.B. and Burgoon, J.K. (1996) Interpersonal Deception Theory, *Communication Theory*, 6, 3, 203-242.
9. Burgoon, J.K. and Buller, D.B. (1994) Interpersonal Deception: III. Effects of Deceit on Perceived Communication and Nonverbal Behavior Dynamics, *Journal of Nonverbal Behavior*, 18, 2, 155-184.
10. Carlson, J.R., George, J.F., Burgoon, J.K., Adkins, M. and White, C.H. (2004) Deception in Computer-Mediated Communication, *Group Decision & Negotiation*, 13, 1, 5-28.
11. Courtis, J.K. (1986) Poor Communication is Alive and Well: A Study of Annual Report Readability, *Canadian Journal of Communication*, 12, 3-4, 1-17.
12. Courtis, J.K. (1987) Fry, Smog, Lix and Rix: Insinuations about Corporate Business Communications, *Journal of Business Communication*, 24, 19-27.
13. Courtis, J.K. (1998) Annual Report Readability Variability: Tests of the Obfuscation Hypothesis, *Accounting Auditing & Accountability Journal*, 11, 4, 459-471.
14. Courtis, J.K. (2004) Corporate Report Obfuscation: Artefact or Phenomenon?, *The British Accounting Review*, 36, 3, 291-312.
15. Dechow, P.M., Sloan, R.G., and Hutton, A.P. (1996) Causes and Consequences of Earnings Manipulation: An Analysis of Firms Subject to Enforcement Actions by the SEC, *Contemporary Accounting Research*, 13, 1, 36.
16. DePaulo, B.M., Lindsay, J.J., Malone, B.E., Muhlenbruck, L., Charlton, K. and Cooper, H. (2003) Cues to Deception, *Psychological Bulletin*, 129, 1, 74-118.
17. Ekman, P. and Friesen, W.V. (1969) Nonverbal Leakage and Clues to Deception, *Psychiatry*, 32, 1, 88-106.
18. Enron Corporate Annual Report (2000).
19. Fowler, R. (1991) *Language in the News: Discourse and Ideology in the Press*, Routledge, London.
20. Fuller, C.M., Biros, D.P. and Wilson, R.L. (2008) Decision Support for Determining Veracity via Linguistic-Based Cues, *Decision Support Systems*, 46, 3, 695-703.
21. Jones, M.J. (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, Autumn, 297-305.
22. Li, F. (2008) Annual Report Readability, Current Earnings, and Earnings Persistence, *Journal of Accounting & Economics*, 45, 221-247.

23. McCornack, S.A. (1992) Information Manipulation Theory, *Communication Monographs*, 59, 1, 1-16.
24. Newman, M.L., Pennebaker, J.W., Berry, D.S. and Richards, J.M. (2003) Lying Words: Predicting Deception from Linguistic Styles, *Personality and Social Psychology Bulletin*, 29, 5, 665-675.
25. Pennebaker, J.W. (2001) Patterns of Natural Language Use: Disclosure, Personality, and Social Integration, *Current Directions in Psychological Science*, 10, 3, 90-93.
26. Rutherford, B. (2003) Obfuscation, Textual Complexity and the Role of Regulated Narrative Accounting Disclosure in Corporate Governance *Journal of Management and Governance*, 7, 2, 187-210.
27. Smith, J.E. and Smith, N.P. (1971) Readability: A Measure of the Performance of the Communication Function of Financial Reporting, *The Accounting Review*, 46, 3, 552-561.
28. Smith, M. and Taffler, R. (1992a) The Chairman's Statement and Corporate Financial Performance, *Accounting & Finance*, 32, 2, 75-90.
29. Smith, M. and Taffler, R. (1992b) Readability and Understandability: Different Measures of the Textual Complexity of Accounting Narrative, *Accounting Auditing & Accountability Journal*, 4, 4, 84-98.
30. Smith, M. and Taffler, R.J. (2000) The Chairman's Statement - A Content Analysis of Discretionary Narrative Disclosures, *Accounting, Auditing & Accountability Journal*, 13, 5, 624-647.
31. Soper, F.J. and Dolphin, J., Robert (1964) Readability and Corporate Annual Reports, *The Accounting Review*, 39, 2, 358-362.
32. Subramanian, R., Insley, R.G. and Blackwell, R.D. (1993) Performance and Readability: A Comparison of Annual Reports of Profitable and Unprofitable Corporations, *Journal of Business Communication*, 30, 49-61.
33. U.S. Securities and Exchange Commission Accounting and Auditing Enforcement Releases, United States Securities and Exchange Commission.
34. U.S. Securities and Exchange Commission (1998) A Plain English Handbook: How to Create Clear SEC Disclosure Documents.
35. Zhou, L., Burgoon, J.K., Nunamaker, J.F., Jr. and Twitchell, D.P. (2004) Automating Linguistics Based Cues for Detecting Deception in Text Based Asynchronous Computer Mediated Communication: An Empirical Investigation, *Group Decision & Negotiation*, 13, 1, 81-106.
36. Zhou, L., Burgoon, J.K. and Twitchell, D.P. (2003) A Longitudinal Analysis of Language Behavior of Deception in E-Mail in: *Intelligence and Security Informatics: First NSF/NIJ Symposium, ISI 2003, Tucson, AZ, USA, June 2-3, 2003 Proceedings,* Hsinchun Chen, R. Moore, D. Zeng and J. Leavitt (eds.), Springer Berlin / Heidelberg, 2003, pp. 102-110.
37. Zhou, L., Twitchell, D.P., Qin, T., Burgoon, J.K. and Nunamaker, J.F., Jr. (2003) An Exploratory Study into Deception Detection in Text-based Computer-Mediated Communication, *Proceedings of the 36th Hawaii International Conference on Systems Sciences (HICSS '03).*