Journal of Forensic & Investigative Accounting
Vol. 6, Issue 1, January - June, 2014

Content Analysis for Detection of Reporting Irregularities: Evidence from Restatements during the SOX-Era

Chuo-Hsuan Lee

Edward J. Lusk

Michael Halperin*

INTRODUCTION AND MOTIVATION

In recent years Content-Based Criteria Analysis (CBCA) has been used to analyze asynchronous communication information embedded in financial documents for detection of irregularities (e.g., Churyk, Lee, and Clinton (2009) [CLC]; Loughran and McDonald, 2011; Humpherys *et al.*, 2011). According to Humpherys *et al.* (2011, p.587), "CBCA is based on the hypothesis that a statement based on fantasy will differ in quality and content from a statement based on actual experience."

Churyk *et al.* (2009) applied such content analysis to analyze the Management Discussion and Analysis [MD&A] section of 10-K reports. They used pre-Sarbanes-Oxley: 2002 [SOX] data comprised of firms, selected in the accrual period 1989-2001, that were issued an AAER (i.e., Accounting and Auditing Enforcement Release) by the SEC during the period of 2000-2003. They found significant differences of language—linguistic coding—used on many dimensions in the MD&A section of 10-K filings between the firms required by the SEC to restate their financial statements and firms not filing such restatements. They offer these linguistic coding differences as a viable method of detection of financial statements with which subsequently the SEC would take issue and require restatement. Humpherys *et al.* (2011) also

* The authors are, respectively, Professor at the State University of New York at Plattsburgh, and Professor at the State University of New York at Plattsburgh and Emeritus Department of Statistics of the Wharton School of the University of Pennsylvania, Philadelphia, and Director Lippincott Library of the Wharton School at University of Pennsylvania.

applied content analysis as one of their methodologies to examine the MD&A section of 10-K reports for the firms involved in the AAER restatements that were issued between 1995 and 2004. They concluded that content analysis is useful in identifying the firms involved in fraudulent financial reports, consistent with the findings of Churyk *et al.* (2009).

The recent work of Churyk *et al.* (2009) and Humpherys *et al.* (2011) provides evidence supporting the usefulness of content analysis as a detection tool. They showed that content analysis is an important screening tool that can signal the likelihood that the SEC will require the firm to file a restated 10-K which may be related to fraudulent reporting. However, the datasets used by Churyk *et al.* (2009) and Humpherys *et al.* (2011) contain firms that have been issued an AAER by the SEC during the period of 2000-2003 and the period of 1995-2004 respectively. In other words, the evidence from their work was mainly based on the data collected during the time period prior to the enactment of the Sarbanes-Oxley Act: 2002 [HR: 3763].

The main purposes of our study are to examine whether or not the content analysis approach is still an effective tool for detection of irregularities or fraud leading to financial statement restatements in the Sarbanes-Oxley [SOX] era, and to determine if the nature of the matching protocol may be a factor affecting the functioning of the CLC version of the CBCA model. Our curiosity was peaked because (1) the SOX has strengthened the regulations for financial reporting, (2) the PCAOB, in its efforts to improve audit quality, has issued stringent requirements of auditing standards, AS 5, mandating more disclosure to be authenticated by the senior managers of the filing organization, and, finally and most importantly, (3) the conclusions of the previous literature of content analysis (e.g., Spathis, 2002; Zhou *et al*., 2004a; Zhou *et al*., 2004b) have become public information available to senior managers. Therefore, it is possible that they could use such knowledge to game these CBCA initiatives which may impact the content analysis screening in the SOX-era.

The key question that we want to address in this study is: As the previous studies succeeded in the detection of irregularities/fraud during the pre-SOX time period, will content analysis be still relevant as a detection tool in the SOX-era? To explore this question, we apply the methodology of Churyk *et al.* (2009) (hereafter, the CLC CA Model) to analyze the differences in the MD&A section of 10-K filings between restatement firms and matched non-restatement firms using the data in the SOX-era: 2002 to 2006. Our major findings are: First, we find essentially identical results to those reported by CLC, strongly suggesting that content analysis when used to analyze the MD&A section of the 10-K in the SOX-era is still effective for detection of irregularities/fraud. This finding is intriguing given the strengthened regulations from the SOX affected through the PCAOB and the increasing attention to content analysis in the literature. Next, as a validation of our accrual, and as a generalization of these replication results, we find that the recent development in the reporting environment shaped by the SOX and the PCAOB has largely expanded the length of the MD&A section of the 10-K issued by both the restatement firms and non-restatement firms. Further, in an exploratory mode, our results were produced under varying matching protocols suggesting that idiosyncratic differences in matching may not be factors that might mask or alter the application of (CBCA) models. Lastly, but not less interestingly, as a suggestion of possible "gaming", we proffer also in an exploratory model that the firms involved in fraud with some understanding of the implications of the public information from the previous content analysis literature will target easy-to-manipulate language cues in order to avoid being detected. In this regard, we find that the firms required by the SEC to restate their financial statements tended to over-use the easy-to-manipulate variable "For example" in the MD&A section which was a reversal from the results reported by CLC.

The implications of our study are important in many respects for practitioners and researchers. First, our evidence confirms the usefulness of using CBCA-linguistic traces in detection of irregularities in the MD&A section of financial reports even after the SOX strengthened MD&A reporting requirements and the information of content analysis has been made publicly available. Second, our replication findings were produced under varying matching protocols suggesting "simple and logical" matching in which the researcher selects the protocol are likely to suffice.[1] Thirdly, as an extension of our results, and to be sure, those of others, it may be necessary for the users of CBCA models to re-evaluate the discriminating power of easy-to-identify individual language cues periodically. Such cues, if they are public information, may offer simple manipulation opportunities to producers of MD&A sections.

LITERATURE REVIEW AND THE CLC CA MODEL

Literature Review

In the SOX-era financial fraud detection has become an important emerging research area (e.g., Zhou *et al.*, 2004a; Lee *et al.*, 2009; Humpherys *et al.*, 2011; Ngai *et al.*, 2011; Glancy and Yadav, 2011). The scandals of *Enron* and *WorldCom* as well as the Madoff-Ponzi scheme heightened investor awareness concerning financial risk due to fraud, and increased the level of public and official oversight over financial reporting. As the gatekeepers of the integrity of financial reporting, auditors have been both enabled as well as challenged because of: (i) the strengthened regulations such as Sarbanes-Oxley: 2002 [HR:3763], (ii) the PCAOB standards now requiring a separate opinion in the assurance report addressing internal control over

[1] Subsequently we detail the two matching protocols that we used.

financial reporting, and (iii) SAS No. 99 requiring auditors to exercise professional skepticism for identification and evaluation of financial fraud risks in the financial statements. For example, Hammersley *et al.* (2011) conducted an experiment to investigate how audit seniors modified an audit program in response to the heightened fraud risk. They find that "audit seniors receiving material weakness information produced audit programs that are no more effective for fraud detection, and are less efficient than those produced by auditors in the control condition" (p.98). Their result implies that audit seniors did respond to the heightened fraud risk but not in an effective way.

According to Humpherys *et al.* (2011), many researchers had, due to the insistence of the PCAOB, perhaps over-focused on improving analytical or statistical procedures to help auditors identify the risk factors related to fraud and red-flag the clients with high risk of fraudulent financial statements (e.g., Loebbecke *et al.*, 1989; Dechow *et al.*, 1995; Beneish, 1999; Bell and Carcello, 2000). The recent development of technology in artificial intelligence and data mining also has provided numerical analytical methods for auditors to detect fraudulent financial statements (e.g., Deshmukh *et al.*, 1997; Lin *et al.*, 2003; Zhang and Zhou, 2004). While most of the previous studies focused on quantitative analysis of financial statements, other researchers used the Content-Based Criteria Analysis (CBCA) to analyze asynchronous communication information embedded in financial statements for fraud detection. According to Humpherys *et al.* (2011, p.587), "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 hypothesis that a statement based on fantasy will differ in quality and content from a statement based on actual experience."

CBCA has been widely used as an effective tool to detect deceit in direct personal communication and text-based asynchronous computer-mediated communication (e.g., Zhou *et al.*, 2004a; Zhou *et al.*, 2004b; Zhou, 2005; Hu *et al.*, 2011). Recently, the application of the CBCA has been extended to analyze the fraudulent 10-K reports (see Churyk *et al.*, 2009; Humpherys *et al.*, 2011; Loughran and McDonald, 2011). Also, some literature has successfully pushed the envelope of the CBCA to examine disclosure documents such as earnings announcements (see Rogers *et al.*, 2011; Feldman *et al*., 2010; Davis *et al.*, 2012) and merger announcements (see Kimbrough and Louis, 2011). For example, Humpherys *et al.* (2011) applied different theories and statistical methods drawn from the field of computational linguistics to investigate the MD&A section of 10-K reports for the firms involved in the AAER restatements that were issued between 1995 and 2004.[2] They found CBCA to be useful in distinguishing between fraudulent firms and non-fraudulent firms based on language cues. They concluded that "the modest success in classification results demonstrates that linguistic models of deception are potentially useful in discriminating deception and managerial fraud in financial statements." [see Humpherys *et al.*, 2011, p. 593]. Loughran and McDonald (2011) obtained a list of 13 red-flagged phrases from a Barron's article and investigate how appearance of these 13 phrases in a firm's 10-K report affects its market performance and fraud allegations. They found that the use of these "problematic" words in the 10-K reports is significantly related to excess filing date return, subsequent return volatility, earnings' forecast dispersion, and fraud allegations. Similarly, Churyk *et al.* (2009) applied the Linguistic Inquiry Word Count Program 2001 (LIWC, 2001) to analyze the MD&A section of 10-K reports for the firms involved in financial restatements between year 1989 and 2001. They found that the CBCA based on LIWC

[2] They did not indicate when the 10-K reports required by the SEC for restatement were issued. Because their AAER restatements were issued between 1995 and 2004 by the SEC, most of their sample firms were collected from the pre-SOX period.

is useful as a tool for detection of fraud, consistent with the finding of Humpherys *et al.* (2011). The details of the CLC CA model used by Churyk *et al.* (2009) are discussed as follows.

The CLC CA Model

Churyk *et al.* (2009) relied on the Linguistic Inquiry Word Count Program: 2001 (LIWC, 2001). According to Churyk *et al.* (2009, p.31), "LIWC parses and identifies parts of speech, and also identifies syntax. LIWC then analyzed the frequencies of the occurrences of language-based cues. The multipart hypothesis was tested by examining these relations as determined by the predefined LIWC 2001 linguistic software dimensions." Based on the findings of prior literature, they considered the following ten variables to be tested in five categories: *Total Words* and *Lexical Diversity* as Standard Linguistic variables; *Colons*, *Semicolons* and *For Example* as Organization Clarity variables; *Positive Emoti*on, O*ptimism and Energy*, and *Anxiety* as Affective and Emotional Processes variables; *Causation* as a Cognitive variable; and *Present Tense* as a Certainty variable. They hypothesized and confirmed that on average the restatement firms relative to the non-restatement firms used: more words, less unique words, less colons, less frequently the term "*For example*", less terms with positive emotions, less terms showing optimism and energy, more terms linked to anxiety, less terms showing causation, and less present tense verbs in their MD&A section of 10-K reports.

HYPOTHESES DEVELOPMENT, CONJECTURE, AND METHODOLOGY

In the SOX-era, the PCAOB through the SOX legislation, the SEC, and the U.S. Department of Justice have created many changes in the financial reporting environment. As a result, management must meet the challenges from (1) increased scrutiny of the corporate operational activities coupled with significant criminal and civil penalties imposed by SOX-sections 404 and 906 for defying internal control and financial reporting mandates, (2) increasing requirements to provide assurance testing not only of the results reported in the three typical financial statements, the Cash Flow, the Income Statement and the Balance Sheet, but also of those reported in the "fourth" Financial Statement (i.e., the Notes to the previously mentioned statements), and (3) the mounting assurance responsibilities relative to the internal control over financial reporting (the so-called COSO section of the report).

The pressures for management to meet the financial reporting challenges during the difficult SOX-era global economy may be seen as providing temptations for management to manipulate their earnings in more creative ways if they desire to "spin" their "engineered" operating results under the tightened regulations. Because content analysis has a history dating back to the mid-1990s, it is natural to ask whether or not management with understanding of the implications from the early content analysis literature such as Zhou *et al.* (2004a) would try to avoid detection in the SOX-era by circumventing the linguistic triggers of content analysis—i.e., gaming the detection protocols. We contend that it is possible that sophisticated management with knowledge of CBCA-content-analysis would try to efface traces in MD&A that may invite red flags from oversight groups. However, in order to do so management must be able to identify all relevant language cues used in the detection driven content analysis, which may not be an easy task to achieve because the nature of the information used by CBCA software makes it complicated for management to stage or manage their language used in creating the MD&A. For example, the LIWC software has various lexicon dimensions and dictionaries, some of which may be difficult for management to maneuver such as *Positive Emotion* as well as *Optimism and Energy* while others may be relatively less complicated for manipulation such as the use of the term "*For example*". Practically speaking, sophisticated managers are certainly capable of procuring CBCA software for self-testing their MD&A. However, this raises a serious legal issue in that such gaming would likely require enlisting individuals conspire to commit fraud by attempting to mask irregularities or fraud at the firm-reporting level. Therefore, the purchase/download and use of the CBCA software would leave an audit trail to incriminate management in this onerous and illegal activity. Accordingly, we expect that content analysis would be still useful in the SOX-era and CLC CA Model should not be diminished in its detection effectiveness. The above rationale leads to our first hypothesis.

Hypothesis 1:

CLC CA Model should not be diminished in its detection effectiveness in the SOX-era.

The variable set that we have selected to test our first hypothesis H1 is taken from the dictionary pertaining to the LIWC: 2001™ version. This dictionary is the one used by Churyk *et al.* (2009). Our seven test variables for H1 are: *Positive Emotion*, *Optimism and Energy*, *Anxiety*, *Causation*, *Present Tense*, *Colons*, and *Semicolons*. We held back the phrase "*For example*" and "*Word count"* as we will test them separately in addressing context-validation testing.[3]

As the next test hypothesis, we argue that the word count in MD&A should increase in the SOX-era because the PCAOB and SOX: [HR 3763] require more disclosure in footnotes and certainly in the MD&A section of the 10-K. To validate our argument, we will examine *Word Count* as a reasonability check on the accrual of firms that form the testing platform of our study and also as an indicator of the strength of the tendency of linguistic encoding for restatement firms. In other words, if the number of words in the MD&A section of the 10-K reports did not

[3] Churyk *et al.* (2009) used another variable, the 10th one in their study call: Lexical Diversity. We did not include this variable as we were not sure if there were other dictionary additions used to create this variable; whereas we were sure that the variables that we selected (i.e., 9 of the 10 of their variables) were exactly the same for our study and their study.

increase, this would cast doubt on the representativeness of our AuditAnalytics™ sample and so the generalization of our replication results. This fact leads to our second hypothesis.

Hypothesis 2:

The recent development in the reporting environment shaped by the SOX and the PCAOB will expand the length of the MD&A section of the 10-K reports issued by both the restatement firms and non-restatement firms.

We now would like to offer a conjecture.[4] As previously mentioned in the development of H1, we argue that the nature of the LIWC software may create some barriers for management to stage or "manage" the language used in the MD&A unless they purchased the LIWC software. Given that purchasing the LIWC software, as discussed above, would leave a download-trail for prosecution, a "safer" way for management to manipulate financial statements to serve a gaming agenda is not to purchase the LIWC software but to use the available implications from the content analysis literature and try to adjust their language cues based on the suggestions of literature. If management does manipulate the language cues in MD&A, our conjecture is that management will tend to manipulate the easy-to-adjust variables such as the term "*For example*". Accordingly, our conjecture is stated as follows.

Conjecture:

The firms required by the SEC to restate their financial statements will tend to over-use the easy-to-manipulate variable: ***For Example****.*

DATA COLLECTION AND SAMPLE SELECTION

Data Collection Processing of the Samples

[4] Lee *et al.* (2009) investigated whether or not the automated deception detection is viable in TAC (text-based, asynchronous, computer-mediated) messages and found that deceivers avoid the defensive targeted features readers associate with lying and include pro-motive-targeted features people associate with truth-telling and conclude that "deceivers construct different messages than truth tellers do and that difference related to pro-motive-targeted features ." (p. 5).

For our study, we downloaded all the firms that had a Revenue/Profit-related restatement as identified in AuditAnalytics™ [from the WRDS™ database] for the time period 2002 to 2006. We selected this five year time period as it was before the *Lehmann Bros. LLP* sub-prime debacle of 2008 but after the SOX became a federal law [HR:3763]. This time period: 2002 to 2006 was in our judgment a relatively "stable" time in the NASDAQ and NYSE which were the markets used as our source of restatement or accrual. We believed that this time period would provide exploratory information on the effect of traversing various event spaces on the functioning of the CLC CA model where they collected data over the time period of 1989 to 2001. These 13 years had many events that likely affected the markets where their accrual firms were listed. For example, the Launch of the WWW circa 1992/3, the creation of the environment that spawned the *dot.coms*, the resulting bubble build-up, the bursting of the same, the defalcations of *Enron*, *WorldCom* and others, and the shocking implosion of Arthur Andersen, LLP. It is certainly reasonable to suppose that such dramatic event-shocks to the markets would have compromised combining the firm accruals over their 13 year longitudinal study. However, remarkably CLC find strong effects. We wondered if these clear linguistic indicators would also be in evidence in the shortened event horizon of the SOX-era.

For our replication, we contacted CLC and received important information on the detailed nature of their protocol. Therefore, we were able to replicate their accrual process of firms and the general nature of their matching.

Specifically, for our study using the Revenue/Profit screen we found that there were 42 firms reported in AuditAnalytics™ for the time period 2002 to 2006 that were requested by the

SEC or that voluntarily filed a 10-K restatement.[5] The year of the restatement was considered as the target year, Y_T,—i.e., the restatement was used as a surrogate for, as CLC note in their paper, fraud. It may be a stretch to identify restatement as being precipitated by a fraud event; therefore we are using restatement as an indication of irregularities [fraud included] that give rise to filing a re-statement. Following the advice of CLC, given the target year, Y_T, we then took the year before the target year as the matching year, Y_M. After we determined the matching firms, we checked to see if at any point past Y_M a restatement was filed for the matching firms for Y_M. If so, we eliminated that firm as match. Also, if there were to have been a re-statement filed for the matching year, Y_M, for the accrual firm (i.e., two restatements in consecutive years), that firm was eliminated as an accrual firm.

After eliminating firms based upon the screening discussed above, due to missing data, or due to the fact that the MD&A was presented in annual reports rather than in the 10-K reports, the final set of accrual firms for which matches were to be sought was 32 firms.[6] This fits well with the number of firms that Churyk, Lee, and Clinton (2009) accrued. They accrued and used 68 firms over thirteen years which is about 5 per year. We accrued 32 firms over five years with about 6 per year.

Given the 32 restatement firms, we identified 114 matching firms not restating, for any reason, their financial statements as the control group. We accrued the 114 matching firms based

---

[5] It is interesting that it is difficult to categorize whether the voluntary re-statement ensued after communication with the SEC. As there was no clear information on if voluntary meant—without any suggestion that the SEC may be going to request a restatement or not—we used the CLC system to just identify firms that filed a restatement and did not try to record WHY they filed a 10-K restatement.

[6] We restricted our accruals to the MD&As from the matching firms for three reasons: 1.) The Information in the MD&A section is about 12 times number of words that are in the annual report based upon a pre-test examination that we did of the two communication modalities! 2.) There is never the same level of detail in the Annual Report as there can be and usually is in the MD&A of the 10-K. 3.) The CLC study protocol used only MD&A of the 10-K reports. Therefore, to maintain consistency and avoid any possibility that the MD&A in the annual reports were somehow different—i.e., the certifying CPA firm sometimes reviews *with comment* the information included in the Annual Report—we eliminated any matches where their MD&A was only in the annual report.

on the following matching criteria. CLC give general information but do not specify in detail how their matching was done. Therefore, we developed our own system of matching. To try to control as much as possible for matching bias in the study the first two authors independently developed matching criteria. This has a further robustness test benefit as there will likely be three systems of matching: the exact CLC matching protocol, unknown to us, and our two. This matching robustness will be mentioned in the results section.

Specifically, the second author, selected as matching variables: *Quick Ratio*, *Market Value of Total Assets*, and *Cash Flows from Operating Activities*. These variables were selected as useful in generating matches because from a previous study of some 30 performance variables for trading firms these three variables loaded on separate factors at values for which the square of the loading was greater than 50%. (Lusk, Halperin, and Heilig, 2009). Then, all of the firms in the same SIC group as the accrual firms for Y_M were downloaded from COMPUSTAT™ [WRDS™]. These potential matches were sorted from highest to lowest by *each* of the three variables. After that the particular accrual firm was located among the sorted firms on each of the three specific variables. For each of the three matching variables: *Quick Ratio*, *Market Value of Total Assets*, and *Cash Flows from Operating Activities*, a set of 20 firms: 10 immediately above <u>and</u> below the location of the accrual firm were selected as possible matches. This gives 60 possible firms as matches—i.e., 20 matches for each of the three variables. The two firms that appeared most frequently in this set of 60 firms were selected as the matching firms. If all 60 firms were unique, a simple random sample was taken.

The first author used *Return on Assets* (*ROA*) and *Total Assets* as the variables for matching. After he downloaded all of the firms in the SIC grouping of the restatement firms, he sorted these firms from highest to lowest by *ROA*. The particular accrual firm was located in the sorted range of *ROA*. Then, five percent of the firms in the sorted download immediately above and below the accrual firm were selected. After repeating the same procedure for *Total Assets*, he examined the selected possible matching firms for both of the variables; a firm was selected to be a match if it appeared in both sets. Two such matches were selected. If two matches were not among the identified firms over the two variables he then incrementally increased the boundaries by five percent on either side for both variables. This process was repeated until two matches were identified or he ran out of the candidate firms in the same industry in which case a random section was made. If the two authors selected the same firm as the match, which happened rarely, then they flipped a coin and the winner kept his match and the loser selected the second best match from his matching system. The reason for identifying duplicates in the matches under the two protocols was to allow another firm to be selected and used to replace the duplicate.

THE RESULTS OF TESTING OUR HYPOTHESES AND CONJECTURE

In the Churyk *et al.* (2009) study, the inferential tests were directional, standard t-tests where the inference was drawn from p-values; we will therefore use the same testing modality. In our study, the number of firms that we identified in the restatement accrual period 2002 to 2006 was 32 for which we identified 114 matches. In comparison, the Churyk *et al.* (2009) study had 68 firms and 118 matches. As these matching protocols created multiple matches, specifically about 3.5 matched per restatement firm for our study, and less for the CLC study, there will be power differences resulting from different sample sizes for detection as well as likely more with-in sample matched-variation for our study. We estimate logically the variation differential to be on the order of 75% which was tested by examining the Welsh-partitions and found to be a reasonable approximation in re-sampled blocked groups over the restatements. [All inferential information was developed using JMP®/SAS©, v.10]. Additionally, the two sample accrual sets of matches developed by the two authors as discussed above were tested to determine if there are any directional differences in the linguistic coding results. The results of the effect differences tested over two accrual protocols suggested that the two authors produced essentially the same results. In this test we used the classifications of the authors as a blocking variable and tested it conservatively with the Median test: Chi^2 version. The result of testing showed that the average p-value was 0.54 and the smallest p-value was 0.11, strongly indicating no classification differences over the study variables for the two sample matching protocols. This simple demonstration is confirmed in that the combined sample used to produce the results of our study generates effect sizes essentially identical to those produced by CLC as displayed in Table 1.

Testing H1 *CLC CA Model should not be diminished in its detection effectiveness in the SOX-era.*

To test whether or not the CLC CA Model remains effective during the SOX-era, we will use parametric t-tests as reported by CLC and additionally re-compute the power corrected for the sample size and variance differentials as discussed above. Specifically, power which speaks to the correctness of the False Positive Error inference is simply computed as [1 – the False Negative Error proportion]. To test the null form of the first hypothesis, we are not only interested in the *direction* and *magnitude* of the effect which, incidentally are NOT modified in the power re-calibration, but also the *power* information. Our results and those of Churyk *et al.* (2009) are reproduced in Table 1.

Table 1 presents clear evidence showing that the CLC CA Model still remained effective during the SOX-era. All the directional effects and almost all the directional magnitudes of our results are remarkably in line with those reported by Churyk *et al.* (2009). Furthermore, the average power re-calibration for our study using a conservative directional FPE of 2.5% is 37%, indicating a convincing corresponding longitudinal similarity to the original study. Finally, the directional p-value for the measured effects using the simple binomial assessment against chance, a conservative test benchmark, is 0.008. All of these indications argue for rejecting the null of H1 that our results were likely to have come from populations where the restatement and their benchmarks were not different—i.e., rejecting the null offers that there are structural differences between the linguistic traces and the benchmarks which is what we observe.

Table 1
Results of Testing Hypothesis 1

	CLC Study Restated Firms: **Means**	CLC Study Matched Firms: **Means**	p-value	Our Study Restated Firms: **Means**	Our Study Matched Firms: **Means**	Adjusted Power Estimates	Result Directional Support? **Yes/No**
Positive Emotions	2.07	2.27	0.005	2.07	2.16	24%	Yes
Optimism and Energy	0.85	0.95	0.02	0.75	0.82	56%	Yes
Anxiety	0.02	0.04	0.01	0.02	0.03	48%	Yes
Causation	1.42	1.52	0.045	1.28	1.29	6%	Yes
Present Tense	2.72	2.93	0.005	2.68	2.76	25%	Yes
Colons	0.10	0.14	.003	0.18	0.18	5%	Yes
Semicolons	0.11	0.15	0.015	0.19	0.40	94%	Yes

Directional p-value for the measured effects using the simple binomial assessment against chance = 0.008

[a] Means are standardized to be comparable to those reported by Churyk *et al.* (2009)

Testing H2 *The recent development in the reporting environment shaped by the SOX and the PCAOB will expand the length of the MD&A section of the 10-K reports issued by both the restatement firms and non-restatement firms.*

As a validation of the representativeness of our accrual protocol, we argue in the second hypothesis H2 that the word count should increase in the SOX-era due to strengthened regulations and reporting requirements. Table 2 presents our results of testing H2.

Table 2
Results of Testing Hypothesis 2

	CLC Results	Our Results	Change from pre-SOX to SOX period
Restated Firms	5,386	11,228	**Increase**
Matched Firms	4,553	10,001	**Increase**
Column P-values	.04	0.31/0.43	

The results in Table 2 indicate that the word count had essentially doubled from the pre-SOX period to the SOX-era period with the respective directional p-values equal to < .0001 for both restated firms and matched firms for both parametric and the Wilcoxon Signed Rank tests. This result supports our second hypothesis and clearly validates our expectation that the SOX scrutiny has resulted in an increase in the details of reporting and so is offered as a validity check of the accrual representativeness and the generalizability of the test results of H1.

Testing our Conjecture: *The firms required by the SEC to restate their financial statements will tend to over-use the easy-to-manipulate variable:* ***For Example****.*

Our conjecture states that management with intention to engage in fraud or produce other irregularities will tend to over-use the easy-to-manipulate variable "*For example*". In Table 3, we demonstrate the testing results concerning our conjecture.

Table 3
Results of testing the *For Example* conjecture

	CLC Results	Our Results	Change from pre-SOX to SOX period
Restated Firms	.003	.0098	**Increase**
Matched Firms	.007	.0035	**Decrease**
Column P-values	.03	0.056/0.006	

Our result in the third column of Table 3 suggests that the frequency (standardized by word count) of using the phrase *For example* for the restated firms is **greater** than that for the matched firms during the SOX-era. Specifically, the null of no difference between restated firms and matched firms is rejected given p-values of 0.056 using the Welsh corrected t-test and 0.006 using the Median test.

In contrast, the CLC result in the second column of Table 3 shows that during the *pre-SOX* period the frequency of using the phrase *For example* for the restated firms is less than that for the matched firms with reported p-value of 0.03. The further comparative analysis between these two columns (i.e., pre-SOX versus SOX-era) indicates that: (1) for the restated firms, the use of *For example* standardized by word count increased from .003 to .0098 which assuming a Wilcoxon Signed Rank test against the CLC result is significant with a two-tailed p-value of 0.08, and (2) for the matched firms, the use of *For example* standardized by word count had declined from .007 to .0035 similarly tested resulting in a two-tailed p-value less than 0.0001. In summary, the comparison across these two periods suggests that on average [standardized for word count] the restated firms had largely increased the use of the phrase *For example* while the matched firms had essentially remained the same in their use of *For example* from pre-SOX to the SOX-era period.[7]

Taken together, the results of Table 3 are consistent with our conjecture and suggest the following possibility. On one hand, management in the restated firms possibly with understanding of the previous content analyses dramatically increased the frequency of the use of the phrase *For example* when preparing their MD&A to avoid inviting scrutiny during the SOX period. On the other hand, management in the matched firms still relied on <u>about</u> <u>the</u> <u>same</u> <u>number</u> <u>of</u> <u>uses</u> of the *For example* phrase—i.e., did not seem to change their behavior re: the MD&A. Therefore, after management in matching firms expanded the length of the MD&A (also, see our discussion for H2), their MD&A presentations showed <u>less</u> frequency *conditional on word count* of the use of the phrase: *For example*. This conjecture certainly merits consideration as a gaming possibility.

SUMMARY AND CONCLUSIONS

Previous studies including Churyk *et al.* (2009) and Humpherys *et al.* (2011) applied a content analysis approach in analyzing the MD&A section of financial reports. They found content analysis to be useful for detection of irregularities or fraud leading to restatement of financial statements. These studies were conducted prior to the enactment of the Sarbanes-Oxley: 2002. In our study, we examined whether or not the content analysis of MD&A is still an

[7] While the matched firms remained the same in their use of the phrase *For example* from Pre-SOX to the SOX-era period, the word count of the MD&A section of their 10-K reports had essentially doubled according to our test result of H2. Therefore, the use of the phrase *For example* standardized by word count declined from 0.007 to 0.0035.

effective tool in the Sarbanes-Oxley era. To explore this question, we applied the methodology of Churyk *et al.* (2009) to analyze the differences between restatement firms and matched non-restatement firms using the data in the SOX period: 2002 to 2006. Our findings are:

1.) The effectiveness of content analysis of MD&A has not been diminished in the SOX-era. This finding is intriguing and speaks to the pervasive and consistent nature of the linguistic traces left in the MD&A section even given strengthened regulations as embodied by SOX and enforced by the PCAOB and the SEC.

2.) There is suggestive exploratory evidence that the nature of matching is not likely to affect or bias the detection results. This opens up the possibility of engaging in creative matching to use as the detection benchmark. This is a critical aspect for forensic monitoring for detection and control as it means that investigative agencies can create idiosyncratic matching protocols that would NOT be known *a-priori* to firms who may consider gaming as a possibility.

3.) The recent development in the reporting environment shaped by the SOX and the PCAOB had largely expanded the length of the MD&A section of the 10-K reports issued by both of the restatement firms and non-restatement firms.

4.) The firms restating their financial statements tend to over-manipulate easy-to-adjust variables such as "*For example*" in the MD&A section of the 10-K. This suggests that the firms involved in fraud and with some understanding of the implications of the publicly available information from the content analyses could target easy-to-manipulate language cues in order to avoid being detected. To simply avoid this tempting gaming possibility, agencies engaged in forensic motoring for control should not reveal monitoring information such as the particular dictionary in use nor the CBCA model applied to conduct the forensic screening. Therefore, we suggest that monitoring agencies (i) vary their linguistic dictionaries, (ii) and/or develop their own dictionaries using the results and datasets of previous studies. This seems particularly crucial given that there are now Internet Links[8] that offer assistance in guiding the preparation of the MD&A sections of 10-K. Also, recently, the *Journal of Accountancy* offered in their Checklist section, a short article titled: *MD&A Reporting Tips* by Senior Editor, Ken Tysiac (Tysiac, 2013). For instance, this short article suggests readers to pay attention to what their competitors and peers are disclosing for "The SEC is going to look at them." (p.20)

OUTLOOK

In summary, our findings—the confirmation of the CLC results through a replication design with extended exploratory features—should underscore the powerful and remarkable result: *Linguistic traces are left when one engages in: defalcation, fraud or constructing irregularities.* This confirms an important opportunity of investigation for auditors, fraud examiners, and researchers who are interested in relying on content analysis for detecting financial statement irregularities.

As our study and that of CLC focus effectively on irregularities—fraud being one aspect of such irregularities—we offer another study begged by our collective results: To identify firms that have been found to have committed fraud, as opposed to non-fraud irregularities, in reporting information in their financial statements. This target of opportunity sample of firms could be examined and compared to a matched benchmark to determine the nature and strength

[8] See: http://www.cpeonline.com/seminar/webcast/webcast---md%2526a-%2526-market-risk-disclosure%3A-a-hands-on-workshop-3

of the linguistic traces left in their MD&As. One would suppose that the linguistic effects would be even more pronounced.

This genera of research puts in play the fascinating and, to be sure, difficult question: What variable set in the Human Information Processing context is instrumental in generating these linguistic coding results—the output of which we observe as linguistic traces left in the MD&A section? In this regard, we offer, in closing for guidance in the development of a research agenda that addresses the HIP context, the poetic one-liner of Sir Walter Scott: "***Oh what a tangled web we weave when first we practice to deceive"***.

REFERENCES

Bell, T.B., and J.V. Carcello. 2000. A decision aid for assessing the likelihood of fraudulent financial reporting. *Auditing: A Journal of Practice & Theory*, 19 (1), 169-184.

Beneish, M. D. 1999. The detection of earnings manipulations. *Financial Analysts Journal*, 55 (5), 24-36.

Churyk, N.T., C.C. Lee, and D.B. Clinton. 2009. Early detection of fraud: evidence from restatements. *Advances is Accounting Behavioral Research*, 12, 25-40.

Davis, A.K., J.M. Diger, and L.M. Sedor. 2012. Beyond the numbers: measuring the information content of earning press release language. *Contemporary Accounting Research*, 29 (3), 845-868.

Dechow, P., R. Sloan, and A. Sweeney. 1995. Detecting earnings management. *The Accounting Review*, 70,193-225.

Deshmukh, A., J. Romine, P., and H. Siegel. 1997. Measurement and combination of red flags to assess the risk of management fraud: A fuzzy set approach. *Managerial Finance,* 23 (6), 35-48.

Feldman, R., S. Govindaraj, J. Livnot, and B. Segal. 2010. Management's tone change, post earnings announcement drift and accruals. *Review of Accounting Studies*, 15, 915-953.

Glancy, F.H., and S.B. Yadav. 2011. A computational model for financial reporting fraud detection. *Decision Support Systems*, 50, 595-601.

Hammersley, J.S., K. M. Johnstone, and K.K. Kadous. 2011. How do audit seniors respond to heightened fraud risk? *Auditing: A Journal of Practice & Theory*, 30 (3), 81-101.

Hu, N., L. Liu, and V. Sambamurthy. 2011. Fraud detection in online consumer reviews. *Decision Support Systems*, 50 (3), 614 -626.

Humpherys, S.L., K.C. Moffitt, M.B. Burns, J.K. Burgoon, and W. F. Felix. 2011. Identification of fraudulent financial statements using linguistic credibility analysis. *Decision Support Systems*, 50, 585-594.

Kimbrough, M.D., and H. Louis. 2011. Voluntary disclosure to influence investor reactions to merger announcements: an examination of conference calls. *The Accounting Review*, 86 (2), 637-667.

Lee, C. C., R.B. Welker, and M.D. Odom. 2009. Features of computer-mediated, text-based messages that support automatable, linguistics-based indicators for deception detection. Journal of Information Systems, 23 (1), 5-24.

Lin, J.W., M.I. Hwang, and J.D. Becker. 2003. A fuzzy neural network for assessing the risk of fraudulent financial reporting. *Managerial Auditing Journal*, 18 (8), 657-665.

Loebbecke, J.K., M.M. Eining, and J.J. Willingham. 1989. Auditors' experience with material irregularities — frequency, nature, and detectability. *Auditing: A Journal of Practice & Theory*, 9 (1), 1-10.

Loughran, T and B. McDonald. 2011. Baron's red flags: do they really work? *Journal of Behavioral Finance*, 12, 90-97.

Lusk, E., M. Halperin and F. Heilig. 2009. Market and Financial Performance as Related to Idiosyncratic Risk and the Effect of Outlier Screening in Market Studies, *Journal of Financial Management and Analysis*, 22:59-69:

Ngai, E., Y. Hu, Y.H. Wong, Y. Chen, and X. Sun. 2011. The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, 50, 559-569.

Rogers, J.L., A.V. Buskirk, and S.L. Zechman. 2011. Disclosure tone and shareholder litigation. *The Accounting Review*, 86 (6), 2155-2183.

Spathis, C.T. 2002. Detecting false financial statements using published data: some evidence from Greece. *Managerial Auditing Journal*, 17 (4), 179-191.

Tysiac, K. 2013. MD&A reporting tips. *Journal of Accountancy*, February, 20.

Zhang, D., and L. Zhou. 2004. Discovering golden nuggets: data mining in financial application. *IEEE Transactions on Systems, Man and Cybernetics*, 34 (4), 513-522.

Zhou, L. 2005. An empirical investigation of deception behavior in instant messaging. *IEEE Transactions on Professional Communication*, 48 (2),147-160.

______, J.K. Burgoon, J.F. Nunamaker Jr., and D.P. Twitchell. 2004a. Automating linguistics based cues for detecting deception in text based asynchronous computer mediated communication: an empirical investigation. *Group Decision and Negotiation*, 13 (1), 81-106.

_____, _____, D.P. Twitchell, T. Qin, and J.F. Nunamaker Jr. 2004b. A comparison of classification methods for predicting deception in computer-mediated communication. *Journal of Management Information Systems*, 20 (4), 139-165.