HIGH-STAKES, REAL-WORLD DECEPTION:

AN EXAMINATION OF THE PROCESS OF

DECEPTION AND DECEPTION DETECTION USING

LINGUISTIC-BASED CUES

By

CHRISTIE M. FULLER

Bachelor of Science
Kansas State University
Manhattan, KS
1998

Master of Business Administration
Fort Hays State University
Hays, KS
2001

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
May, 2008

HIGH-STAKES, REAL-WORLD DECEPTION:

AN EXAMINATION OF THE PROCESS OF

DECEPTION AND DECEPTION DETECTION USING

LINGUISTIC-BASED CUES

Dissertation Approved:

Rick L. Wilson

Dissertation Adviser

David P. Biros

Dursun Delen

Carol Johnson

A. Gordon Emslie

Dean of the Graduate College

ACKNOWLEDGEMENTS

I wish to extend sincere thanks to Dr. Rick Wilson, my committee chair, for his support and guidance throughout my years at OSU. I also extend my gratitude to Dr. David Biros for his guidance and for introducing me to deception research.

Sincere thanks are also extended to Dr. Dursun Delen and Dr. Carol Johnson, members of my doctoral committee, for their suggestions. I wish to thank Dr. Judee Burgoon for her input during the early stages of this study. I also wish to express my appreciation to the other MSIS faculty, who have supported me in a variety of ways over the last five years. To my peers in the MSIS doctoral program, I appreciate your support and wish you every success as you complete this process.

I could not have completed this work without the support of my friends and family. My mother, Carolyn Fuller, was my biggest champion throughout this process. Special thanks also go to my sister, Natalie Fuller, and my father, Larry Fuller, for their encouragement.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

Deception has previously been defined as “a message knowingly transmitted by a sender to foster a false belief or conclusion by the receiver”(Buller & Burgoon, 1996). Methods of deception detection have existed for thousands of years. As noted by one author, for as long as people have been lying, people have been trying to detect deception (Ford, 2006). Despite this history of attempting to detect deception, humans have not proven to be very capable at this task, as most are not able to detect deception at a rate better than chance (Vrij, Edward, Roberts, & Bull, 2000). Methods that could assist humans with the task of deception detection are intrusive, subjective, or fail to achieve acceptable accuracy levels on a consistent basis. They may also require extensive user training.

Accurate, non-invasive, user-friendly methods are needed to address the shortcomings of existing deception detection methods. Improved methods of deception detection are particularly important to those who must detect lies in the usual course of their work, such as security personnel, human resource managers, among others. Automated classification methods have been introduced into text-based deception research as one possible alternative to previous methods (Zhou, Burgoon, Nunamaker, & Twitchell, 2004; Zhou, Burgoon, Twitchell, Qin, & Nunamaker, 2004).

Background

Prevailing theories of deception include Interpersonal Deception Theory (Buller & Burgoon, 1996), Information Manipulation Theory (McCornack, 1992), Four Factor Theory (Zuckerman & Driver, 1987), Ekman's Clues to Deceit, (Ekman, 1985; Ekman & Friesen, 1969) and Reality Monitoring (Johnson & Raye, 1981). Additionally, known cues to deception have recently been summarized in the self-presentational perspective of deception (DePaulo et al., 2003).

When discussing deception, the terms 'lying' and 'deceit' are often used interchangeably (Buller & Burgoon, 1996; Ekman, 1985; Grover, 1993; Vrij & Mann, 2004). The concept of deception also includes the choice to lie. Those who unintentionally provide false information are not considered to be engaging in deception.

Common to theories of deception is a focus on cues to deception. These indicators, or cues, may be divided into three classes: nonverbal, paraverbal, and verbal (Sporer & Schwandt, 2006). Nonverbal cues are those such as eye contact and body movements. Paraverbal cues are vocal cues that accompany speech, such as voice pitch. The third category is verbal content cues, such as pronoun usage and verbal immediacy.

A subset of verbal content cues, linguistic-based cues, has been defined to describe those cues that can be operationalized with general linguistics knowledge (Zhou, Burgoon, Nunamaker et al., 2004). Linguistic-based cues are relatively content independent and lend themselves to automated analysis. In early deception research, nonverbal cues received more attention (Berrien & Huntington, 1943; Cutrow, Parks, Lucas, & Thomas, 1972; Ekman & Friesen, 1969; Ekman & Friesen, 1972; Ekman & Friesen, 1974). Over time, nonverbal, paraverbal and verbal content cues to deception have been studied, with a recent meta-analysis listing 158 of these cues (DePaulo et al., 2003). This list includes indicators belonging to all three classes of cues.

Recent studies have begun to focus on deception detection using linguistic-based cues and their possible utility in automated deception detection (Zhou, Burgoon, Nunamaker et al., 2004; Zhou, Burgoon, & Twitchell, 2003). However, the foundation for this technique has not been empirically validated using traditional methods, such as factor analysis. Additionally, these cues have not been validated in a 'high-stakes' real-world context.

Problem statement

Humans are not very accurate lie detectors. A recent study summarizing results of over 23,000 subjects found the average accuracy in detecting deception to be 54 percent (Bond and DePaulo, 2006). Several alternate methods exist for deception detection including the polygraph, Statement Validity Analysis, and Reality Monitoring (Vrij et al., 2000). A summary of existing deception detection methods and the type of cues they use is shown below in Table I.

Automated deception detection using linguistic-based cues is a method that holds promise (Zhou, Burgoon, Nunamaker et al., 2004). It is not invasive, does not require complex training and is especially relevant given the rise in text-based communication in everyday life (Zhou, Burgoon et al., 2003) and the difficulty with which people recognize verbal forms of deceit. Automating the analysis also provides less subjective results.

Table I

Cue Classes and Related Deception Detection Methods

Cue Class	Applicable Methods	Drawbacks of Methods
Nonverbal	Polygraph	Invasive, specialized equipment required, extensive training required, subjective results
	Brain Fingerprinting	Invasive, specialized equipment required, proprietary technology
Paraverbal	Voice Stress Analysis	Inaccurate, subjective results
Verbal Content	Scientific Content Analysis	Extensive training required, subjective results
	Content Based Criteria Analysis	Extensive training required, subjective results, not appropriate for use with suspect statements
	Automated Text-Based Deception Detection	New, not previously tested with real-world data
Mixed	Behavioral Analysis Interview	Extensive training required, subjective results, inaccurate

One study showed that people lie in 14 percent of emails and 21 percent of instant messages (Hancock, Thom-Santelli, & Ritchie, 2004). Yet another study by George and Keane (2006) examining deceptive resumes found that respondents identified less than a third of the deceptions in text. This suggests a need for research in deception that analyzes text. The technique presented here is not the only technique for analyzing the veracity of text. However, in contrast to other methods, this is an automated technique, which should allow it to readily analyze large data sets.

While the sample here is written text produced as part of the investigation of crimes, the technique should be applicable to other forms of text. For example, electronically produced text such as email (such as the Enron corpus), web pages, or blogs could be analyzed with the method employed here. It might also be used with transcribed text of oral communications.

Most studies in deception detection, regardless of approach, use student subjects in experimental settings (Vrij & Mann, 2001b). A recent meta-analysis of 120 studies showed 101 used student subjects. Only four of these studies involved situations where the subjects were not given instructions as to whether they should lie, but subjects did so on their own (DePaulo et al., 2003). It is interesting to note that there is evidence that behavior differs between those who choose to lie and those who lie at the direction of an experimenter (Feeley & deTurck, 1998).

Therefore, studies utilizing real-world samples of subjects who either chose to be truthful or deceptive may contribute more deeply to the understanding of deception. Previous studies in linguistic analysis of deception have relied on 'mock lies' (Zhou, Burgoon, Nunamaker et al., 2004; Zhou, Burgoon, Twitchell et al., 2004) or very small samples (n=18) (Twitchell, Biros, Forsgren, Burgoon, & Nunamaker Jr, 2005), leaving analyzing real-world data largely unexplored. A past polygraph study comparing mock crime and actual field data found significant differences in results between the samples (Pollina, Dollins, Senter, Krapohl, & Ryan, 2004), underscoring the need for real-world examination of deception.

Previous researchers have noted result inconsistencies across deception studies. Differing subjects of lies--either a subject's attitude or feelings versus description of actual event--may contribute to the mixed results seen across previous research (DePaulo et al., 2003; Ford, 2006; Miller & Stiff, 1993; Sporer & Schwandt, 2006). It is thought that the cues that emerge may vary with the subject of the lie, though the set of cues related to any particular situation has yet to be empirically identified.

Another possible explanation for the inconsistency in findings across studies is whether actual or perceived cues to deception were studied (Bond & DePaulo, 2006; Vrij, 2000). "The question of identifying particular cues, qualities, or cue combinations that lead an observer to infer dishonesty is of course quite distinct from the question of identifying the particular cues that really do signal duplicity…" (DePaulo, Zuckerman, & Rosenthal, 1980). It is the second statement that drives this study.

The Krauss, Geller, and Olsen study (1976) illustrates the need to distinguish between actual and perceived cues to deception. In this study, subjects were asked to describe the information they used to detect lies, they were observed to see what information they actually used to detect lies, and behavior was analyzed to see what cues differentiated truthful and deceptive communication. The cues reported by the subjects, the cues the subjects were observed to be using, and the actual cues to deception were each different. Vrij (2000) also found differences in actual and perceived indicators of deception.

Research Contributions

As described above, there is a need for methods of detecting deception in text. A framework of linguistic constructs has been proposed for this purpose (Burgoon, Qin, & Twitchell, 2006; Zhou, Burgoon, Nunamaker et al., 2004; Zhou, Burgoon, Twitchell et al., 2004). This framework has been used successfully to distinguish truthful and deceptive messages, but it has not been empirically validated. The framework relies, in part, on deception theories, but which theories are valid in the context of the current study is not known. Combining this framework with a review of deception theory, this study aimed to answer the following research question:

- What are the appropriate constructs for use in studying deception in text?

Using this framework as a starting point, this study refined a set of constructs for studying text-based deception. Further, using appropriate statistical methods, the study empirically validated these defined constructs.

In addition to validating constructs for use in studying text-based deception, this study aimed to further the understanding of how people deceive when using written or other text-based communication methods. As past results have been inconsistent across studies, and these cues have not been studied extensively in text, or using 'real-world' data, (DePaulo et al., 2003; Zuckerman & Driver, 1987), it is unknown precisely what the moderating effect will be on verbal cues in real-world data. To address these issues, this study determined which linguistic based cues distinguish truthful from deceptive messages in a high-stakes environment. A closely related issue is the impact of incident severity on the production of cues in truthful and deceptive messages. The investigation of these issues was guided by the following research questions, specifically for text-based environments:

- Which linguistic-based cues distinguish truthful and deceptive subjects in a high stakes environment?
- How does severity impact cue intensity?

Determining which theoretically-based constructs and cues distinguish truthful and deceptive messages contributes to the understanding of how deception takes place or how deceptive messages, as a group, can be expected to differ from the set of truthful messages. A classification model, using linguistic-based cues, can aid in determining the veracity of individual messages. Using a variety of cue sets and models, the study showed that classification of real-world, text-based deception data can be done accurately with a parsimonious cue set. As the upper limit on sample sizes for studies of this type will likely continue to be restricted due to the difficulty of establishing ‘ground truth’, identifying the best cues is pertinent. To guide this portion of the study, the following research question was developed:

- Can the veracity of individual messages be accurately determined using linguistic-based cues?

In summary, this study aimed to achieve four primary objectives: First, a set of linguistic constructs was validated for use in text-based, high-stakes deception detection research. Next, the study identified linguistic-based cues that accurately distinguish truthful and deceptive messages with a factual subject in a high-stakes environment. Third, the severity of the situation impact in cue importance within this environment was examined, and finally, an accurate classification model was constructed using linguistic-based cues in this real-world, high-stakes environment.

This chapter has introduced the topic of the research, outlined the need for the study, and introduced the research contributions resulting from the study. Chapter Two will review the relevant literature. Chapter Three will describe the constructs studied and describe the methodology and results related to validating these constructs. Chapter Four will describe the hypotheses tested, the related methodology, and the results of testing the hypotheses. Chapter Five will detail the classification portion of the analysis. Chapter Six will discuss these findings and conclude the dissertation.

CHAPTER II

REVIEW OF LITERATURE

Several methods of detecting deception exist, each with various advantages and disadvantages. This review begins with a discussion of these methods. Deception theory may provide a basis for new deception detection methods, including automated deception detection using linguistic-based cues, also termed automated-text based deception detection. A discussion of relevant theory follows the review of current deception detection methods.

Deception Detection

Methods for detecting deception have existed for thousands of years, some less scientific than others. Over three thousand years ago in China, suspects were forced to place dry rice in their mouths. If the rice was still dry when they spit it out, the suspect was thought to be lying. In medieval times, deception detection was known to involve walking on hot coals or being dunked in water (Ford, 2006). Over time, more sophisticated methods of deception detection have developed.

Though deception detection is not a new practice, humans have not proven to be very capable at this task. A synthesis of over 23,000 subjects showed that human performance at the task of deception detection is just slightly better than chance (Bond & DePaulo, 2006). Ekman and O'Sullivan (1991) have found just 15 out of 13,000 people who can detect deception with 80 percent accuracy. Previous studies have shown that professional lie catchers, with few exceptions, are generally no better than college students or the general public at detecting deception. A summary of eight studies of professional lie catchers, such as police officers or customs officers, shows overall accuracy levels ranging from 49 to 64 percent (Vrij, 2000). Further, it has been suggested that professionals may be more difficult to teach to detect deceit (Vrij, 2000), perhaps due to a reluctance to abandon old habits and beliefs. Other studies of varying populations have also found training to be unsuccessful (Akehurst, Bull, Vrij, & Kohnken, 2004; Biros et al., 2005). This suggests that deception detection training is a daunting task regardless of trainee background. Even where training has shown success, the improvement due to training has been very limited. A summary of training showed that trained observers were 57 percent accurate, whereas their untrained counterparts were 54 percent accurate (Vrij, 2000).

Secret service agents are one of the few groups that have been found to be significantly better than chance at deception detection. Psychiatrists who have an interest in deception also are better than most. Other groups of professionals for which lie detection may be a necessary or desirable skill, such as judges, police officers and regular psychiatrists, have not performed as well (Ekman & O'Sullivan, 1991; Ekman, O'Sullivan, & Frank, 1999; Grubin & Madsen, 2005).

These results report overall accuracy, or the accuracy of detecting truthful and deceptive messages combined. When separating the results for detecting lies and the truth, the results may be even worse. It appears that humans may be reasonably good at detecting the truth, with truth accuracy of 70-80 percent. The accuracy at correctly detecting lies is only 35-40 percent (Feeley & Young, 1998). This may reflect the notion of 'truth bias' or the idea that receivers are more likely to judge communications to be truthful than deceptive (Levine, Kim, Park, & Hughes, 2006; Park & Levine, 2001; Vrij, 2000). The poor record of humans at lie detection may be attributable to a belief in global signs of lying, which may not exist, and/or incorrect beliefs or lack of knowledge about the cues that actually point to deception in particular circumstances (Feeley & deTurck, 1995; Fiedler & Walka, 1993; Grubin & Madsen, 2005; Mann, Vrij, & Bull, 2004; Vrij, 2000). To aid humans in the task of deception detection, several methods have been developed, including those for analyzing text. These include Automated Text-Based Deception Detection, Scientific Content Analysis, Statement Validity Analysis, and the Behavioral Analysis Interview.

Automated Text-Based Deception Detection

Linguistic analysis tools have been introduced as a possible aid in deception detection that may address some of the drawbacks of other methods (Zhou, Burgoon, Nunamaker et al., 2004). Of 158 recently listed cues to deception, (DePaulo et al., 2003) approximately 50 could potentially be used in analysis of text. Within this subset of cues, some are defined rather ambiguously and are not strictly cues for text. One example of such a cue is unusual contents. While a human could read a piece of text and make a determination whether the information was relevant and fit within the context of what is being described, accomplishing this task with text-processing tools is quite difficult, and perhaps not even possible (DePaulo et al., 2003). Good candidates for automated analysis are those cues that can be analyzed objectively and can be defined in a manner relatively independent of the content of the text (Zhou, Burgoon, Nunamaker et al., 2004). There are two prevailing tools currently being used to analyze deception in verbal communication using just some of these linguistic-based cues: Agent 99 Analyzer and LIWC.

Agent 99 Analyzer

At the University of Arizona, a tool labeled Agent 99 has been developed for use in automated deception detection in a variety of forms, including text (Cao, Crews, Lin, Burgoon, & Nunamaker, 2003; Zhou, Twitchell, Qin, Burgoon, & Nunamaker, 2003). Within Agent 99, the tool developed for deception detection in text has been labeled Agent 99 Analyzer (A99A). This tool relies heavily on Generalized Architecture for Text Engineering (GATE) (Cunningham, 2002; Cunningham et al., 2005), for text processing. Waikato Environment for Knowledge Analysis (WEKA) (Witten & Frank, 2000) is used for classification based on the initial text processing steps.

Utilizing cues belonging to a variety of categories (quantity, complexity, uncertainty, non-immediacy, expressivity, diversity, informality, specificity, and affect), the use of linguistic-based cues in deception has been investigated (Burgoon et al., 2006; Qin, Burgoon, Blair, & Nunamaker Jr, 2005). Studies have found several cues that significantly differ between truthful and deceptive messages using both desert survival and mock theft scenarios (Burgoon, Blair, Qin, & Nunamaker, 2003; Zhou, Burgoon, Nunamaker et al., 2004; Zhou, Twitchell et al., 2003).

In addition to examining significant differences between truthful and deceptive message groups, A99A has also been used for classification studies. In one study, with a sample of 94 messages of student subjects, overall accuracy in classifying statements as truthful or deceptive ranged from 57.4 percent using a decision tree to 80.2 percent using an artificial neural network (Zhou, Burgoon, Twitchell et al., 2004). The performance of the classifier was increased when a subset of the most relevant cues was used. Another study with a small real-world data set of 18 messages achieved accuracy of 72 percent (Twitchell et al., 2005). This system has also been used to study the effect of modality in deception (Qin et al., 2005), finding that differences between truthful and deceptive messages remained fairly consistent across modalities. Overall, about thirty cues have been used in the Agent 99 Analyzer studies, with up to 22 cues used in any individual study.

LIWC

Linguistic Inquiry and Word Count (LIWC) processes text based on four main dimensions: standard linguistic dimensions, psychological processes, relativity, and personal concerns (Pennebaker & Francis, 2001). Within each of these dimensions, a number of variables are represented. For example, the psychological processes dimension contains variables representing affective and emotional processes, cognitive processes, sensory and perceptual processes, and social processes. In total, the default dictionary serves as the basis for 74 output variables. LIWC was initially created to identify basic cognitive and emotional dimensions and has since been expanded and refined.

Newman, Pennebaker, Berry, and Richards (2003) proposed that the language dimensions of self-references, negative emotions, and cognitive complexity could be associated with deception. The use of motion and exclusive words were proposed as indicators of cognitive complexity. The study found that third person pronouns were also a predictor of deception. They used LIWC to extract the variables described above and then classified text using logistic regression. The overall accuracy in this study was 61 percent.

Based on the work of Newman et al. (2003), Bond and Lee (2005) used LIWC to code the statements of prisoners. Utilizing the variables from the previously described study by Newman et al, a classification accuracy of 69.1 percent was achieved using the prisoner statements. In addition to the categories studied by Newman et al., Bond and Lee also used LIWC to code Reality Monitoring (RM) Terms. Bond and Lee had an overall accuracy rate of 71.1 percent using logistic regression to classify statements based on the RM terms.

Hancock and colleagues (2004) have also examined the use of automated linguistic analysis in deception. Their research, which draws on Interpersonal Deception Theory (Buller & Burgoon, 1996) and the self-presentation perspective (DePaulo et al., 2003; Vrij, 2000), hypothesized differences in word counts, pronoun usage, words related to feelings and senses and exclusive words between deceptive and truthful communications. The study used LIWC to analyze eight variables in the four categories described above. Deceptive senders used more words, more third person pronouns such as “he”, “she” and “they”, and more sensory terms than truthful senders.

This work was later expanded to include evaluation of three additional variables: negative emotions, causation terms (such as “because” and “effect”), and question marks (Hancock, Curry, Goorha, & Woodworth, 2005). Motivated senders used significantly more causation terms than unmotivated senders. Significant effects were not found for the other variables. Classification models were not implemented in the studies by Hancock and colleagues.

The results of the above studies show that methods incorporating automated linguistic analysis show promise in studying deception and deception detection. Linguistic analysis techniques may overcome many of the limitations (invasiveness, inconsistent accuracy, the need for extensive user training, time consuming procedures, and the required presence of a trained examiner) of previously introduced methods, Other techniques that can evaluate text include SCAN, CBCA, and BAI.

Scientific Content Analysis

Scientific Content Analysis (SCAN) is a statement analysis procedure developed for use in criminal investigations. SCAN relies on criteria such as pronoun usage, spontaneous corrections, emotions, and connection phrases in analyzing transcripts or written statements. These criteria are not unique to SCAN and many are quite similar to those used in Criterion Based Content Analysis. Based on the limited published results available, the technique appears to work reasonably well in classifying statements as truthful or deceptive. It has been noted that the technique may not work when the subject is discussing multiple issues (Driscoll, 1994). This technique was created by Avinoam Sapir, a former Israeli police lieutenant, based on years of experience interrogating subjects and is not theoretically based (Lesce, 1990; Porter & Yuille, 1996). The technique's accuracy has been compared to that of the polygraph, though specific accuracy rates have not been reported

Statement Validity Analysis and Content Based Criteria Analysis

Statement Validity Analysis (SVA) is a technique for analyzing the verbal content of statements. It is made up of three components, one of which is Content Based Criteria Analysis (CBCA). SVA was originally developed for determining the veracity of the testimony of children in sexual abuse cases, but has since been more widely applied to other types of cases and to adult subjects. CBCA, the SVA component which receives the most attention, involves analyzing a statement according to 19 criteria. CBCA is based on the Undeutsch hypothesis, which posits that statements derived from memories from actual events differ from statements that are based on fantasy (Undeutsch & Yuille, 1989). Beyond this conjecture, the technique lacks theoretical foundation (Sporer, 1997; Vrij, 2000). While the full list of criteria includes 19 items, a subset of 14 criteria are sometimes used, as the full list might be applicable only when the technique is used for its original purpose (Vrij, 2000). The results of past studies analyzing the statements of adults have shown that the technique's accuracy may vary widely, with reported accuracy ranging from 55 to 90 percent.

There is also a truth bias associated with CBCA, as results show that the technique works better for detecting truths than lies (Vrij, 2000). This is particularly problematic in the context of crime investigation, since the focus is identifying deceptive statements accurately. A recent study showed that there may be issues to address in achieving inter-rater reliability when using this technique (Godert, 2005), though if raters are trained properly, the technique can be more effective. If aspects of the technique can be automated, this particular issue can be somewhat lessened. However, the nature of the technique is that the criteria are subjective (Vrij, 2000), so there will be limits to how much inter-rater reliability can be improved. The subjective nature of CBCA also limits its potential for automation.

Further, despite the fact that CBCA was developed to be used as just one of three parts of SVA, in 2000, one researcher remarked that not a single SVA study had been published. There are no formal rules for determining whether a statement analyzed using CBCA is truthful or deceptive (Vrij, 2000), such as how many criteria must be present or how the criteria should be weighted. It has also been suggested that due to its design, CBCA may not be appropriate for use with suspect statements (Vrij, 2005).

Behavioral Analysis Interview

The Behavioral Analysis Interview (BAI) is a method of deception detection that relies on observing suspect verbal and nonverbal behavior during a structured interview (Horvath, Jayne, & Buckley, 1994). In one study, four judges trained in using the technique reviewed 60 tapes of actual suspects that were interviewed using BAI.

Overall, raters correctly identified truthful suspects with 78 percent accuracy and deceptive statements with 66 percent accuracy. No conclusion was drawn 15.5 percent of the time. A more recent study showed found that suspects' behavior in BAI interviews was not consistent with the types of behaviors predicted by the technique. The updated study did not assess the ability of rater's to distinguish between truthful and deceptive suspects (Vrij, Mann, & Fisher, 2006). Like CBCA and SCAN, this technique relies on a trained rater's assessments of various criteria.

Polygraph

While the focus of this study is deception detection in text, for the sake of comparison, it is worth mentioning other prevalent methods of deception detection. The polygraph is perhaps the most well-known tool that may be used to assist humans with the task of deception detection. The device measures changes in physiological activity and an examiner makes a veracity determination based on these changes (Vrij, 2000). The polygraph was invented in 1917 by William Moulton Marston (Ford, 2006), though he was not the first to experiment with pulse and blood pressure as measures of deceit. The device was expanded to include heart rate, blood pressure, respiratory rate, and galvanic skin response in 1932.

The polygraph has been shown to be one of the most accurate lie detection methods. One report gives accuracy rates between 72 and 91 percent in field studies (National Research Council, 2002). Despite its popularity and apparent accuracy, the polygraph is not without significant drawbacks. First, the results of the polygraph examination are heavily dependent on the examiner (Sporer, 1997; Vrij, 2000). Second, extensive training is required to obtain certification to administer a polygraph examination. There may also be practical limitations to the use of the polygraph as both the appropriate equipment and a trained examiner must be available. This test is also considered intrusive (Twitchell, Jensen, Burgoon, & Nunamaker, 2004). During the test, several sensors are attached to the subject's body, pneumatic tubes are put around the chest and stomach, and a blood pressure cuff is placed around the subject's arm (Vrij, 2000). Finally, the test may be time-consuming. The examiner must arrive at the necessary location, the instrument must be calibrated, the exam must proceed using the required format, and then the results must be evaluated. The test can only be used to analyze responses to ‘yes’ or ‘no’ questions.

Voice-Stress Analysis

The voice stress analyzer was introduced in the 1970s and touted as a possible replacement of the polygraph (Rice, 1978). The voice stress analyzer measures psycho-physiological responses of the suspect. Like the BAI technique, the interrogation must be properly structured and the machine’s results must be carefully interpreted. Unlike the polygraph, it can be used without the subject’s knowledge, though such practice is not without controversy. The accuracy of voice stress analyzers is reported to range from chance level (Gamer, Rill, Vossel, & Godert, 2006; Vrij, 2005) to about equal to that of the polygraph. These machines are fundamentally designed to detect stress, not lies and, like the polygraph, are heavily dependent on the skill of the operator when used as lie detectors. Despite its initial promise, the voice stress analyzer has failed to gain scientific acceptance (Ford, 2006; Hollien & Harnsberger, 2006; Hopkins, Benincasa, Ratley, & Grieco, 2005).

Methods of the Future

Brain Fingerprinting has been offered as yet another alternative to the polygraph. However, the technique is patented, so while the results with this technique have been promising, only limited studies have been published. In one study, the technique correctly classified all six subjects. This technique may prove to be highly accurate, but involves even more time and preparation than the polygraph (Ford, 2006). Similarly, other groups of researchers are working on finding structural areas of the brain associated with lying.

Though there are numerous methods of deception detection, automated text-based deception detection has the potential to accurately determine veracity using minimal resources. Previous results using this technique have been encouraging, though the theoretical foundation for this method has not been validated. The basis of automated lie detection should be those cues that actually indicate deception versus those that observers might perceive to indicate deception. The origin of these cues should be those identified by deception theory describing the process of how people actually deceive. Several theories of deception have been developed, though not specifically for use in text. These theories are described next.

Theories of Deception

Knapp et al Hypotheses

In the 1970's, Knapp and colleagues made a set of predictions as part of an effort to define deception as a communication construct (Knapp, Hart, & Dennis, 1974). This work hypothesized that deceivers would be more uncertain, vague, nervous, reticent, dependent, and unpleasant. Though this work does not rise to the level of a theory, later works share some elements with this study. Further, this study coded a number of verbal behaviors, including some linguistic-based cues, and was one of the first to do so.

Ekman's Clues to Deceit

Ekman (1985, 1992, 2001) describes two kinds of clues to deceit: leakage and deception clues. Leakage describes the mistakes deceivers make that reveal the truth. Deception clues reveal that deception is taking place but do not reveal the truth. From deception clues we can determine whether someone is lying; from leakage, we can determine what it is a person is lying about. Most cues that have been studied would be considered deception clues (DePaulo et al., 2003). Whether cues appear as leakage or deception clues, thinking and feeling aspects to deception are described that may drive the production of cues. The primary feelings, or emotions related to deception are fear, guilt and duping delight. Liars may fear getting caught, feel guilty about lying or experience excitement associated with the challenge of getting away with the lie. Even though a liar may try to conceal these feelings, they may not be able to control all expression of clues associated with them. Thinking cues include inconsistencies, appearing over-rehearsed, and speaking slowly.

Four Factor Theory

Zuckerman et al. (1987) defined four factors involved in deception that can influence behavior: attempted control, arousal, felt emotion, and cognitive processing. According to this theory, deceivers will try to control their behavior to prevent disclosure of deception which will then reveal cues to deception such as behavior that appears planned, rehearsed or lacking in spontaneity. The behavior of the deceiver may also seem overexaggerated. This theory is similar to Ekman's thinking cues to deception (Ekman, 1985). The four factor theory also suggests that deceit will be associated with physiological arousal. The deceit is believed to cause changes in several non-verbal behaviors such as pupil dilation and eye blinks. It may also increase speech errors. Deception also is associated with affect, specifically negative affects such as guilt and anxiety. This echoes Ekman's feeling cues (Ekman, 1985). The Four-Factor Theory suggests a cognitive component to deception. It is believed that deception is more difficult than telling the truth. This complexity will lead to identifiable changes in the behavior of the subject such as more frequent hesitations, and a decrease in frequency of illustrators. It is noted that some of the behaviors associated with cognitive complexity may also be related to arousal and that it may not be possible to isolate exact causal antecedents.

Reality Monitoring

Reality monitoring theorizes that memories based on actual experiences and memories based on imagined events are distinct on several dimensions (Johnson & Raye, 1981). While not originally developed as a theory of deception, the theory has been extended to this context (Frank & Ekman, 1997; Vrij & Mann, 2001a, 2001b). Truthful accounts are expected to share characteristics with memories based on actual experience, and deceptive accounts are expected to share characteristics with imagined events (Vrij et al., 2000). Specifically, real memories will contain more perceptual information, contextual information, and affective information. Imagined events are expected to include more cognitive operations and be more vague (Vrij et al., 2000). Sporer (1997) developed a set of reality monitoring criteria to be used to distinguish truthful and deceptive communications: clarity, perceptual information, spatial information, temporal information, affect, reconstructability of the story, realism, and cognitive operations. Vrij (2000) later provided a review of ten studies using these criteria.

Clarity refers to whether the statement is clear and vivid. One study showed truthful statements to have greater objective clarity. Perceptual information refers largely to whether the statement includes sensory information, visual details and details of physical sensations. Seven of the reviewed studies showed this criterion to be greater in truthful accounts. The third criterion, spatial information, has produced mixed results. This criterion refers to information about locations, and the arrangement of people and objects. Temporal information has been found to be greater in truthful accounts. This type of information involves statements that include information about when the event happened. Affect, or details about the subject's feelings during the event, has also shown mixed results. Story reconstructability is expected to be greater in truthful accounts based on past studies, as is realism or the extent to which the study is plausible and realistic. Seven of the studies found no relationship between cognitive operations and deception.

It has been noted (DePaulo et al., 2003) that most people do not create lies entirely from scratch, but derive them largely based on experienced events, so reality monitoring may be most applicable in those situations where deceivers are creating their tales entirely from scratch and truthtellers are relaying facts, as was generally the case in the studies reviewed by Vrij (2000). This perspective may be less applicable in a situation where a deceiver may send a message that is simply a modification of actual events.

A recent review of reality monitoring research notes no known studies where "real statements by real witnesses are analyzed" (Masip, Sporer, Garrido, & Herrero, 2005), though its accuracy in classifying various types of statements using either discriminant analysis or logistic regression has been shown to be up to 85 percent using laboratory data. The similarity between the reality monitoring criteria and those used in CBCA has also been noted (Sporer, 1997, 2004). An important distinction is that reality monitoring is a theory that has been applied to deception and deception detection, whereas CBCA is a deception detection method that is only very loosely, if at all, grounded in theory.

Self-Presentational Perspective of Deception

In earlier work, Depaulo (1992) described the self-presentational perspective of nonverbal communication. A recent deception meta-analysis expands upon this perspective by organizing the combined cues to deception into five categories representing nonverbal, verbal and paraverbal communication. The first category suggests that liars are less forthcoming than truthtellers. According to this category, liars should provide shorter and less detailed responses. Deceivers may also seem reticent. The second category predicts that liars will tell less compelling tales. That is, their messages will include more discrepancies, be less engaging, more passive, uncertain, and non-immediate. The third category predicts that liars will be less positive and pleasant. Fourth, liars are predicted to be more tense.

The final category of this perspective on deception predicts that liars will include fewer ordinary imperfections and unusual contents within their messages. This last category largely includes those cues that are part of CBCA. The self-presentational perspective is largely based on the pretext that most lies that are told are ‘everyday lies’ (Sporer & Schwandt, 2006). While there is overlap in the predictions of the self-presentational perspective with prior theories of deception, this perspective may be most applicable for use with everyday lies.

Information Manipulation Theory

Information Manipulation Theory (IMT) proposes that deceptive messages violate the conversational maxims of quality, quantity, relation and manner (McCornack, 1992). These conversational maxims were proposed by Grice as guidelines for effective and efficient use of language. In deceptive communication, the quantity of the information may be manipulated simply by altering the amount of information that is presented. Quality manipulations would be represented by what might be considered stereotypical deceptive messages. These manipulations involve deliberate distortions or fabrication of information. Relation violations of conversational maxims occur when the relevance of information is manipulated. For example, a subject may fail to directly answer a question. The final way that IMT suggests that messages are manipulated is through manner. Here, information is conveyed in an ambiguous fashion or will lack clarity.

Interpersonal Deception Theory

Interpersonal Deception Theory (IDT) views deception as an interactive form of communication, merging the principles of deception with the principles of interpersonal communication (Buller & Burgoon, 1996). Though originally developed for study of deception in richer media, such as face-to-face communication, later work has suggested that it is applicable for studying most forms of communication, due to its view of deception as a strategic undertaking, which is not restricted to nonverbal environments

(Zhou, Burgoon, Nunamaker et al., 2004). According to the authors of IDT, communication includes both strategic and nonstrategic behaviors (Buller & Burgoon, 1996). Within the context of IDT, strategic behavior refers to large-scale plans and intentions, not necessarily to specific routines or tactics. Related to this strategic behavior, deceivers may engage in information management, image management, and behavior management during interpersonal communication.

Information management is a key aspect of IDT, reflecting how deceivers control information with the goal of creating credible message (Burgoon, Buller, Guerroro, Afifi, & Feldman, 1996). According to information management, deceivers alter their messages along the following dimensions: veracity, completeness, directness/relevance, clarity, and personalization. Image management includes attempts to maximize credibility of the sender, such as managing one's demeanor to appear competent and trustworthy. Behavior management reflects additional efforts to prevent leakage by controlling behavior that might expose deception. IDT's notion of behavior management is similar to Zuckerman's dimension of control. While information management, image management, and behavior management are all considered part of strategic behavior by IDT, information management is most closely related to verbal behavior (Burgoon et al., 1996).

Non-strategic, or inadvertent behaviors reflect unintentional, unconscious behavior. Non-strategic behaviors have also been labeled leakage (Ekman & Friesen, 1969). IDT suggests that deceivers will unintentionally display arousal, negative affect, and noninvolvement. This is consistent with Zuckerman's view that deception would influence changes in affect and arousal (Buller & Burgoon, 1996; Zuckerman & Driver, 1987).

Information Management and Information Manipulation Theory

As noted above, according to IDT, information management is one strategy used by deceivers when trying to create credible messages. There are five main dimensions to information management. Again these are veracity, completeness, directness/relevance, clarity, and personalization. Several of these dimensions correspond to the four dimensions of quality, quantity, relation, and manner that are outlined by IMT. In a previous work, Burgoon and colleagues have described how the dimensions of information management and IMT are related (Burgoon et al., 1996).

The first dimension of information management is veridicality. This may also be conceptualized as truthfulness, honesty, veracity, or message fidelity and is quite similar to the IMT maxim of quality. This dimension describes stereotypical notions of truth or how the truth is expected to appear. Completeness is similar to the IMT maxim of quantity or whether the speaker has provided as much information as the circumstance requires. The Directness/relevance dimension of information management as described by IDT is similar to the IMT maxim of relation. This dimension describes the extent to which the message is relevant to the context and circumstance.

Clarity is similar to the IMT maxim of manner. Clarity describes speech that should be clear, comprehensible, and concise. Deception may be signaled by communications that are vague and ambiguous. Personalization, also termed disassociation or verbal nonimmediacy, describes whether a person's own thoughts, opinions and feelings are reflected by the information. This dimension is linked with the construct of nonimmediacy (Wiener & Mehrabian, 1968). Nonimmediate language is used to distance or disassociate the speaker from the message. Increased modifiers, which have been suggested as an indicator of clarity, have also been suggested as a sign of nonimmediacy, along with generalizations, shifting time and place of events, etc.

Though most of the theories of deception were developed separately, they share many common elements. These commonalities are summarized in Table II below. From this summary, it can be seen that the existing literature suggests nine dimensions or constructs that can be used to describe deception in text-base communication, though not all may be amenable to automated analysis. The fit of these constructs for both the text-based environment and automated analysis will be assessed in a later section. In addition to examining messages to determine their veracity using appropriate constructs, this study will also look at the impact of severity, or high-stakes situations, on the content of the message

Table II

Common Elements of Deception Theories

Element	Original Name	Theory
Veridicality	Veridicality	IDT
	Quality	IMT
Completeness	Completeness	IDT
	Quantity	IMT
	Perceptual/Contextual Information	Reality Monitoring
	Reticent	Knapp et al.
	Less Forth Coming	Self-Presentational Perspective
Directness/ relevance	Directness/relevance	IDT
	Relation	IMT
	Uncertain	Knapp et al.
Clarity	Clarity	IDT
	Manner	IMT
	Vague	Knapp et al.
Immediacy	Personalization	IDT
	Dependent	Knapp et al.
	Less compelling tales	Self-Presentational Perspective
Arousal	Arousal	IDT, Four-Factor Theory
	More Tense	Self-Presentational Perspective
	Nervous	Knapp et al.
Affect	Negative Affect	IDT
	Feeling Cues	Ekman's Clues to Deceit
	Felt Emotion	Four-Factor Theory
	Affective Information	Reality Monitoring
	Unpleasant	Knapp et al.
	Less Positive and Pleasant	Self-Presentational Perspective
Control	Control	Four-Factor Theory
	Behavioral/Image Management	IDT
Cognitive Processing	Cognitive Processing	Four-Factor Theory
	Cognitive Information	Reality Monitoring
	Thinking Cues	Ekman's Clues to Deceit

High Stakes Deception

Much deception research has focused on everyday lies (DePaulo, Kirkendol, Kashy, Wyer, & Epstein, 1996; DePaulo et al., 2003; Frank & Feeley, 2003; Vrij et al., 2000). While everyday lies may comprise most of the lies people tell (DePaulo et al., 1996), understanding more serious, or high-stakes, lies, and detecting those lies has been deemed important (DePaulo et al., 2003; Frank & Feeley, 2003; Kohnken, 1985). High-stakes situations are those in which the subject has something to gain or lose by being judged truthful or deceptive (Frank & Ekman, 1997).

Past research has found that the cues that are significant under conditions of low motivation are different than those that are significant under conditions of high motivation (Zuckerman & Driver, 1987). In a low motivation state, Zuckerman & Driver found eight cues to be significantly different between truthful and deceptive subjects. In a high motivation condition, ten cues were significantly different between the groups. Only three cues (pupil dilation, blinking, and speech hesitations) had significant differences between deceptive and truthful conditions for both conditions. For eight cues, there was a significant difference in the level of the cue between high- and low-motivation conditions, where only one such difference would have been expected by chance. Thus it appears that motivation does impact cue production. In order to understand how subjects react when the consequences are of importance and the subjects are, therefore, presumably motivated, researchers have introduced the concept of ‘high-stakes’ deception (Frank & Ekman, 1997; Vrij & Mann, 2001a, 2001b).

There have been limited studies that have studied real-world high-stakes deception. These include a study in which police officers studied tapes of the requests for help from the public of five subjects later identified to be guilty in the case in question (Vrij & Mann, 2001a). In this study, the police officers were 50 percent accurate in identifying deception, a rate equal to that of chance. The police officers did slightly better (64 percent) when watching tapes of confessed murderers in a study which purports to be the first published study of a real-world high-stakes situation (Vrij, 2000, 2005; Vrij & Mann, 2001b).

In a related study, the frequencies of six non-verbal behaviors were coded from tapes of sixteen suspects, including: hand movements, shifting positions, foot and leg movements, gestures, self-manipulations, and hand/finger movements. Three of the nonverbal behaviors were shown to be indicators of deceptive behavior. The other cues were inconsistent across the sample, with about half the deceivers showing an increase in the behavior and half the truthtellers showing an increase in the behavior (Mann, Vrij, & Bull, 2002). A99A has also been used to evaluate a small sample of real-world statements with promising results (Sporer, 2004). These studies of real-world samples are particularly important, as it has been questioned whether even 'mock crime' paradigms can provide understanding into how deception occurs naturally (Ekman, 1985; Pollina et al., 2004).

In addition to the studies described above using real-world data, some high-stakes studies, using laboratory data, have used cues that can be analyzed in text in an automated manner, and therefore may provide some foundation for the current study. These include studies based on a single detection methodology, such as CBCA (Akehurst et al., 2004; Godert, Gamer, Rill, & Vossel, 2005); or studies that have utilized cues from multiple theories and methods (Adams, 2002; Davis, Markus, Walters, Vorus, & Connors, 2005; Porter & Yuille, 1995; Vrij, Edward, & Bull, 2001; Zhou, Burgoon, Twitchell et al., 2004; Zhou, Twitchell et al., 2003).

Several studies of high-stakes deception have included only nonverbal cues (Berrien & Huntington, 1943; Bradley & Janisse, 1981; Gamer et al., 2006; Hocking & Leathers, 1980; Mann et al., 2002; Stromwall, Hartwig, & Granhag, 2006; Vrij, 1993; Vrij, 1995; Vrij, Semin, & Bull, 1996) or have only evaluated deception detection accuracy or perceptions of deception rather than evaluating actual cues to deception (Feeley & Young, 2000; Frank & Ekman, 1997, 2004; Granhag & Stromwall, 2001; Kraut & Poe, 1980; Lakhani & Taylor, 2003; Mann et al., 2004; Meissner & Kassin, 2002; Vrij & Mann, 2001a). Only limited studies of high-stakes deception, using both real-world data and text analysis, have been conducted. More studies in this area can contribute to an understanding of the cues to deception in this context.

This chapter has described existing deception detection methods and deception theories. The constructs that are common to these theories were also summarized. Additionally, the need for high-stakes deception research was summarized. The following chapter will detail the development of a set of constructs for studying this domain as well as the validation of these constructs.

CHAPTER III

TEXT-BASED DECEPTION CONSTRUCTS

Construct Development

A review of the literature suggested several constructs (see Table II) that might be appropriate to guide the study of deception in text. These constructs were compared to a previously developed set of constructs, the Zhou/Burgoon framework, which were developed for this purpose. In addition, prior construct validation attempts provided insight useful in further development of this set of constructs. Using this information, the final set of constructs to be studied was defined.

<u>Prior Construct Validation Attempts</u>

A key part of this study was the refinement and validation of a framework for deceptive text-based communication; therefore, it seems pertinent to note prior efforts to define the behavioral dimensions or constructs related to deception. Vrij and colleagues (1996) conducted principle components analysis on six nonverbal behaviors (self-manipulations, shifting positions, hand and finger movements, foot and leg movements, gestures, and head movements). The result of the analysis, using an orthogonal rotation, was three factors, each including two variables. The factors were labeled nervous behavior, subtle movements, and supportive behavior. Given the limited number variables, or cues, included in this study, the applicability of the results may be limited. Further, it may be difficult to support the argument of orthogonal factors for describing deceptive behavior, limiting the relevance of this analysis to the current context. It is specifically noted that the Information Management dimensions of IDT are conceptualized as non-independent (Burgoon et al., 1996). The three factors suggested by Vrij and colleagues seem only applicable to describing nonverbal behavior and seem unlikely to extend to a more general description of deceptive behavior.

Sporer and colleagues have conducted factor analysis, also with orthogonal rotation, on data analyzed using CBCA and Reality Monitoring criteria in several studies (Sporer, 2004). The results have not been entirely consistent, though sufficient commonalities have emerged to suggest five dimensions: Logical consistency/realism, clarity/vividness, quantity of details and contextual embedding, feelings and thoughts, and verbal/non-verbal interactions. The interactions dimension consists only of criteria from CBCA, while the other dimensions consist of criteria from both CBCA and Reality Monitoring. There seems to be little relation between the first dimension, logical consistency/realism and the constructs suggested by the review of deception theory, as summarized in Table II. Several theories, including IMT and IDT, suggest constructs consistent with the clarity/vividness and quantity of details dimensions.

Feelings and thoughts have also been previously suggested as important aspects of deception, though they are usually discussed as separate aspects of deception (Buller & Burgoon, 1996; DePaulo et al., 2003; Ekman, 1985; Zuckerman & Driver, 1987). Though the factor analysis conducted in these studies is based in part on the CBCA, which is not theoretically based, the results are consistent, in part, with deception theory, and were useful in refinement of the Zhou/Burgoon constructs. However, like the study of nonverbal behavior (Vrij et al., 1996), the factor analysis relied on orthogonal rotation. As described previously, an oblique rotation may be more appropriate for the theories reviewed in this study.

Zhou/Burgoon Linguistic-Based Cues Framework

In operationalizing the constructs of the revised framework, the research of Burgoon, Zhou and colleagues (Burgoon et al., 2006; Zhou, Burgoon, Nunamaker et al., 2004; Zhou, Burgoon et al., 2003; Zhou, Burgoon, Twitchell et al., 2004; Zhou, Twitchell et al., 2003) in automated deception detection using linguistic-based cues serves as a strong base. The constructs of the Zhou/Burgoon framework (referred to as the Zhou/Burgoon cues or framework from here forward) were subjected to construct validation along with the revised constructs in order to determine which of these competing frameworks is more appropriate to the current high-stakes context.

For this study, the Zhou/Burgoon framework was reviewed for consistency and completeness relative to those constructs suggested by deception theory, as summarized in Table II. Recently, Burgoon, Qin, and Twitchell (2006) published an updated version of the framework, consisting of eight categories, as shown in Table III.

Table III

Zhou/Burgoon Linguistic-Based Cues Framework

Construct	Variables
Quantity	Words, Verbs, Sentences
Specificity	1^{st} person pronouns, 2^{nd} person pronouns, 3^{rd} person pronouns, other references, modifiers, sensory ratio and number of sensory details
Affect	Affect , Imagery, Pleasantness
Diversity	Lexical Diversity, Content word diversity, Redundancy
Complexity	Average sentence Length, Average word length, pausality.
Uncertainty	Modal Verbs
Nonimmediacy	Passive voice
Activation	Emotiveness, activation

The Zhou/Burgoon framework summarized linguistic-based cues using the following categories: Quantity, specificity, affect, diversity, complexity, uncertainty, nonimmediacy, and activation. Quantity suggests reticence by deceivers, leading to manipulations in the number of words and sentences. This is conceptualized similar to the completeness and quantity dimensions of IDT and IMT, respectively. Specificity implies that deceivers will manipulate the level of details present. While this dimension has some similarity to quantity in that it reflects the amount of information included in the message, quantity refers more generally to length details while specificity reflects type of details, such as described by reality monitoring (Zhou, Burgoon, Nunamaker et al., 2004) or CBCA.

Affect is defined as "A feeling or emotion as distinguished from cognition, thought, or action." (*American heritage dictionary*, 1991). The affect construct has been used to represent the emotions present in the message or language (Whissell, 1989; Zhou, Burgoon, Nunamaker et al., 2004). Diversity is viewed as an extension of completeness, or quantity. This reflects the extent to which an appropriate amount of detail has been included in the message. While the concept of diversity is somewhat similar to both specificity and quantity, it is meant to reflect the level of detail in general, rather than specific types of details, such as spatial or temporal information.

The category of complexity was selected for the framework based on a previous study of newspaper credibility (Burgoon, Burgoon, & Wilkinson, 1981). It primarily refers to how simple the message or language is or is not. Uncertainty refers to evasive or ambiguous language used to avoid giving direct or relevant answers. Nonimmediacy is used in messages to avoid taking responsibility for or claiming ownership of the message. Activation attempts to capture the expressivity of the language used (Fuller, Biros, Adkins et al., 2006; Fuller, Biros, Twitchell, Burgoon, & Adkins, 2006; Zhou, Burgoon, Nunamaker et al., 2004).

Revised Constructs

Based on existing literature and the Zhou/Burgoon framework, a set of revised constructs were developed for use in studying deception in text. Though the Zhou/Burgoon framework was a useful starting point, it did not cover all pertinent aspects of deception, as not all dimensions identified in Table II are included within the framework. Further, the Zhou/Burgoon framework constructs have not been empirically validated. Without this validation, there cannot be any assurance that the cues that are being measured are indicators of the related constructs (Pedhazur & Schmelkin, 1991). Therefore, the revised set of constructs were meant to more fully describe deception and provide guidance for measuring the appropriate cues in the current study and future research. The proposed constructs can be divided into two components. The first set of constructs reflects the difference between the group of truthful messages and the group of deceptive messages. Seven constructs were proposed belonging to this group. The second component includes the construct of severity which was expected to impact the content of the messages and specifically the intensity of cues in the context of high-stakes, factual message production.

Deception

The summary of deception theory (see Table II) was reviewed for constructs that could be operationalized using linguistic-based cues. This information was used to guide refinement of the Zhou/Burgoon framework. All constructs in a refined framework should be theoretically supported, as well as amenable to measurement by automated methods.

This review of the Zhou/Burgoon framework, which utilized both the review of prior theory and previous construct validation attempts, suggested the addition of new constructs, omission of constructs that are not theoretically supported, and improved measurement of the constructs. The first construct considered was *Completeness*. This construct represented whether the message includes an appropriate amount of information. This may refer both the amount of detail present in the message and the length of the message. The Zhou/Burgoon framework separates this construct into those of Quantity and Specificity. A previous factor analysis showed that Quantity of Details and Contextual Embedding should be part of the same construct (Sporer, 2004). In the revised framework, the length of the message and the amount and type of detail in the message were separated. To be consistent with the Zhou/Burgoon framework, these constructs were labeled *Quantity* to represent message length and *Specificity* to represent amount and type of details. The Specificity construct subsumed the Specificity and Diversity constructs of the Zhou/Burgoon linguistic-based cues framework.

Directness related to the relevance of the information to the context and circumstance. It includes the level of uncertainty, or strength and firmness of the passage (Knapp et al., 1974). Uncertainty may reflect attempts to avoid giving relevant answers (Fuller, Biros, Twitchell et al., 2006). In the Zhou/Burgoon framework, this concept was referred to as *uncertainty*.

The next construct suggested for the revised construct set was *clarity*. Clarity describe the degree to which messages were clear and comprehensible (Burgoon et al., 1996). Comprehension expressed the ease of understanding a message (Burgoon et al., 1981). Messages may lack clarity by demonstrating vague and ambiguous language. The factor analysis of RM and CBCA criteria suggested a factor termed Clarity/Vividness. The category in the Zhou/Burgoon framework perhaps shared greatest conceptual similarity with the construct of clarity was the complexity category.

Immediacy, or as it may alternatively be termed non-immediacy was considered to be related to veracity by several previous descriptions (Buller & Burgoon, 1996; DePaulo et al., 2003; Knapp et al., 1974). Immediacy described whether the message includes attempts to disassociate oneself from the events described. Language that implies claiming responsibility for the message content was also included within the definition of this construct (Fuller, Biros, Twitchell et al., 2006). *Affect* has frequently been included in deception theory (Buller & Burgoon, 1996; DePaulo et al., 2003; Ekman, 1985; Zuckerman & Driver, 1987). Terms used to refer to this construct include: affect or affective information, feeling cues, and felt emotion. A previous factor analysis included criteria including psychological processes, cognitive operations, and emotions in a factor labeled feelings and thoughts (Sporer, 2004). Despite this result, to be consistent with the larger set of literature (Buller & Burgoon, 1996; DePaulo et al., 2003; Ekman, 1985; Zuckerman & Driver, 1987), affect was separated from constructs representing thinking or cognitive operations.

The final deception construct is that of *cognitive processing*. It is thought that the difficulty involved in being deceptive differs from that of being truthful (Vrij, 2000). This discrepancy should lead to identifiable changes in behavior and related cue production (Zuckerman & Driver, 1987). Recently, researchers conducting automated text analysis have successfully integrated cognitive processing-related variables (Bond & Lee, 2005; Newman et al., 2003). This construct is not represented in the Zhou/Burgoon framework.

Severity

The constructs described above were intended to describe elements that may differ between deceptive and truthful messages. An additional aspect of this study was the examination of deception in a 'high-stakes' environment, which has been described as situations in which the subject has something to gain or lose by having his or her message judged to be truthful or deceptive (Frank & Feeley, 2003). Within the high-stakes context of this study, not all messages had the same potential consequences, and therefore there were expected to be differences within the group of deceptive messages and within the group of truthful messages. In order to examine the impact of varying consequences, this study utilized the concept of severity. This concept was defined in terms of punishment related with involvement in the incident. This construct was studied in relation to the set of constructs that represent various aspects of deceptive messages. The full list of constructs to be examined is summarized in Table IV.

Table IV

Summary of Constructs to be Studied

Construct	Theoretical Foundation	Brief Description
		Deception Constructs
Quantity	IDT, IMT, Self-Presentational Perspective	Length of message
Specificity	IDT, RM	Amount and type of details in the message
Uncertainty	IDT, IMT	Relevance, directness, and certainty of message
Clarity	IDT, IMT	Message clarity and comprehensibility
Immediacy	IDT, Self-Presentational Perspective	Attempts to disassociate oneself from the events described
Affect	IDT, Ekman’s Clues to Deceit, Four-Factor Theory, RM, Self-Presentational Perspective	Emotions present in the message
Cognitive Processing	Four-Factor Theory, RM, Ekman’s Clues to Deceit	Increased or decreased cognitive processing and cognitive information present in the message related to veracity
		Impact of Severity
Severity	Frank & Ekman, 1997, Vrij, 2000, Depaulo et al., 2003	Consequences of being involved in incident described

There were a few constructs suggested by reviewing the literature that were not included in the revised framework. *Veridicality* was considered for inclusion in the revised framework. This construct refers to the overall truthfulness or appearance of truthfulness of the message (McCornack, 1992). This construct represents typical beliefs regarding honesty. As it captures the overall truthfulness of the message (Buller & Burgoon, 1996), it could not be logically separated from the deception variable. Therefore veracity was not included in the revised framework. Zuckerman et al (1987) proposed *control* as one of the four aspects of deception. This construct is not well-defined for the environment of automated analysis. Therefore, control was not included in the revised framework.

Arousal is also suggested as a possible construct in previous deception literature (Buller, Burgoon, Buslig, & Roiger, 1996; DePaulo et al., 2003; Knapp et al., 1974; Zuckerman, DePaulo, & Rosenthal, 1981). Previous research has proposed that it may not be possible to separate the cues of arousal from those of cognitive complexity or affect (DePaulo et al., 2003; Zuckerman et al., 1981; Zuckerman & Driver, 1987). To achieve consistency with the Zhou/Burgoon framework, this construct was not included in the revised framework.

The Zhou/Burgoon framework includes an *activation* category. Activation, also termed expressivity may be considered as one component of affect (Whissell, 1989). While research in newspaper style related expressivity, or emotiveness, to trustworthiness (Burgoon et al., 1981), this category has not otherwise been theoretically supported, except as it may have some relation with affect. While it may have some predictive value, due to lack of theoretical support, it was not included in the revised framework. Next, the data sample used for the dissertation, and the methods used to validate the constructs are described. Then, the results of construct validation are detailed.

Construct Validation Methodology

Data sample

The sample for the study was a subset of those who completed a statement, officially known as a report 1168, at two military bases from January 2002 to December 2006. This data sample was used for construct validation, and all subsequent analyses.

Person-of-interest statements are official reports written by a subject or witness in an investigation. The process of recording an incident statement from a person of interest is as follows: the investigators typically have the person-of-interest come into the office where they have the option to write the statement or type it into a computer. The statements were all recorded on AF Form 1168. If the person-of-interest is simply a witness and not actively involved in the case, his or her statement could be recorded in the field. All statements were written in the presence of law enforcement personnel. If a person is a suspect he or she was read both the Miranda rights and Article 32 of the Uniform Code Military Justice prior to making a statement. Base law enforcement personnel reviewed cases to find those in which the statement could be identified as truthful or deceptive. Table V describes the criteria used to identify statements.

Table V

Criteria for Determining Statement Veracity

Statement Type	Criteria
Deceptive	1. The subject later recanted the statement and recorded another statement, but was not charged with making a false official statement. 2. The subject was charged with making a false official statement. 3. Other evidence in the case showed that the statement could not be true. 4. An impartial witness, such as security force personnel, gave a statement substantially contradicting the subject's statement.
Truthful	1. Evidence in the case or result of the case corroborated the statement. 2. Statement is given by law enforcement personnel witnessing the incident. Law enforcement personnel are assumed to be impartial witnesses who would make every attempt to give reliable accounts.

This sample provided the opportunity to examine deception and its detection in a real-world, high-stakes context. As previously noted, most studies have been conducted in experimental settings using student subjects (Vrij & Mann, 2001; Depaulo et al., 2003). A need for research using serious, or high-stakes, lies has also been identified (DePaulo et al., 2003; Frank & Feeley, 2003; Kohnken, 1985). Due to these circumstances, all available statements that base personnel could confidently identify as truthful or deceptive were collected from the military bases. Of the over 370 statements gathered to date, many more truthful than deceptive statements were received.

Procedures to prepare the statements for analysis, including transcription of the original statements were prepared by the team of researchers involved in the project. This process included removing identifying information, typing the statement exactly how it was written while coding for any anomalies that could not be transcribed directly, and saving the transcribed statement with a specified file name that captured additional information about the statement. The complete procedure is shown in Appendix B.

Construct Validation

The first major goal of this study was to validate a framework of constructs for use in research of text-based deception and its detection. To meet this goal, based on theory and past literature, a set of revised constructs were proposed and a set of linguistic-based cues were identified to measure each of these constructs. Confirmatory factor analysis was used to empirically validate the proposed constructs. To establish whether or not the revised set of constructs were superior to the previously unvalidated Zhou/Burgoon framework, it was also analyzed using confirmatory factor analysis.

Measurement of Constructs

The proposed constructs of this study had to be validated before hypotheses could be appropriately developed. This validation required the identification of appropriate cues to measure each construct; therefore this section was developed prior to hypothesis development. Many of the cues used as indicators of the defined constructs were retained from the Zhou/Burgoon framework. Where improved measures were available, they were substituted for existing measures.

Quantity was the first construct that has been defined for the revised set of constructs. In the Zhou/Burgoon framework and previous studies (Burgoon et al., 2006; Qin et al., 2005), number of words, number of verbs, and number of sentences were used to measure quantity. As this construct was not substantially redefined, this construct continued to be measured by these three cues.

Specificity was defined for this study to generally reflect the amount of details and type of details included in the message. This definition was used in order to be consistent with the Zhou/Burgoon framework (Burgoon et al., 2006) and findings of a previous factor analysis (Sporer, 2004). Spatial information, temporal information and sensory information are appropriate cues to represent different types of details that might be present in a statement. These cues have previously been used in the study of deception in text (Zhou, Burgoon, Nunamaker et al., 2004; Zhou, Burgoon et al., 2003).

Previously, specificity was measured by 1st person pronouns, 2nd person pronouns, 3rd person pronouns, other references, modifiers, sensory ratio and number of sensory details. To more closely fit the current definition, specificity does not include measures of pronoun usage, as these measures may be more closely related to the revised construct of immediacy. Previously, lexical and content word diversity were used to measure the amount of details, though in a separate category of the framework (Burgoon et al., 2006; Zhou, Burgoon, Nunamaker et al., 2004). Lexical diversity is measured as the ratio of different words or terms to total terms.

Content word diversity is measured by the number of content words divided by the total number of words. It has been shown that lexical diversity is dependent on text length. An alternate measure, bilogarithmic type-token ratio, has been developed to deal with this problem (Kohnken, 1985). The bilogarithmic type-token-ratio was used as a substitute for lexical diversity in the revised constructs.

Uncertainty was only measured with one variable--modal verbs--in the Zhou/Burgoon framework. This can be problematic (Pedhazur & Schmelkin, 1991) as the sources of systematic and nonsystematic variance cannot be identified in single-indicator constructs. Therefore, additional variables were included to measure this construct in the revised framework. A certainty dictionary was available within the LIWC system (Pennebaker & Francis, 2001). Tentative constructions have previously been associated with the uncertainty in deception (Knapp et al., 1974). A variable termed Tentative is included within LIWC, and was used to measure this aspect of uncertainty. Passive voice terms have also been considered to be an indicator of uncertainty (Knapp et al., 1974; Zhou, Burgoon, Twitchell et al., 2004) and were used as such in this study. A final proposed indicator of uncertainty was generalizing terms (Zhou, Burgoon, Nunamaker et al., 2004).

Clarity has previously been measured by average word length, average sentence length, and pausality (Burgoon et al., 2006; Zhou, Burgoon, Twitchell et al., 2004). As defined in the revised framework, comprehensibility is part of the construct of clarity. Previously, factor analysis has shown that comprehension is correlated with readability, redundancy, sentence length and complexity. In this work, complexity was defined as the ratio of characters to words and syllables to words. Redundancy is a measure of function words (articles, prepositions, and conjunctions) per sentence. Readability measures emphasize word and sentence length (Burgoon et al., 1981). Causation words, such as *because* or *effect*, are believed to add a level of concreteness to the message or make it less vague (Hancock et al., 2005) contributing to the clarity of the message. Based on these previous studies, average word and sentence length, redundancy, causation words and complexity ratio were used to measure the clarity construct.

Immediacy describes attempts to associate oneself with a message or to claim ownership of its content. Immediacy is indicated by pronoun usage. Self-oriented terms might be used to associate oneself with the message. Other-oriented pronouns may signal attempts to distance oneself from the message (Hancock et al., 2005). Self-oriented terms include first-person pronouns. Other-oriented terms include second and third-person pronouns (Hancock et al., 2005; Zhou, Burgoon, Nunamaker et al., 2004).

The definition of *affect* in the revised framework is consistent with previous forms of this construct (Zhou, Burgoon, Twitchell et al., 2004). The Whissell Dictionary of Affect in Language uses three variables to measure this construct: activation, imagery, and pleasantness (Whissell, 1989). This dictionary includes a total of 8742 words and has been tested for reliability and validity. This dictionary has previously been integrated into deception studies (Burgoon et al., 2006; Fuller, Biros, Adkins et al., 2006; Fuller, Biros, Twitchell et al., 2006; Zhou, Burgoon, Twitchell et al., 2004). In one study, activation was moved to a separate category containing this variable and expressivity (Burgoon et al., 2006). For measuring the revised constructs, activation was included in the affect construct, in order to be consistent with the Dictionary of Affect.

Cognitive Processing is not a construct in the Zhou/Burgoon framework, but has been studied previously in deception studies. The use of motion verbs and exclusive words have been associated with deception (Newman et al., 2003). Additionally, a dictionary of cognitive processing terms was used as an indicator for this construct based on previous work (Pennebaker & Francis, 2001).

Severity was determined from the subject of the statement. As described above, severity was used in this study to capture the consequences of the event described. Deception researchers have established a need to study serious lies, though few studies have studied linguistic-based cues in this environment. Further, measures have not previously been established to capture this concept.

Severity was introduced in this study on a somewhat exploratory basis as an attempt to capture how differing levels of severity or stakes impacted the production of linguistic-based cues. All types of incidents within the sample were given a rating between one and five by law enforcement personnel, with one being the least severe and five representing the highest severity. To achieve interrater-reliability, the ratings were made by three experienced law enforcement officials, with eight to fourteen years of experience. To determine the statement severity, the subject of each statement was identified, and marked with the corresponding rating. The revised constructs and their related measures are summarized in Table VI.

To extract the cues from the statements, GATE and LIWC were used. In A99A, General Architecture for Text Engineering (GATE) (Cunningham, 2002; Cunningham et al., 2005) is used to extract cues from text and Waikato Environment for Knowledge Analysis (WEKA) (Witten & Frank, 2000) to build classification models using the extracted cues. GATE has successfully been used in the past for extracting linguistic-based cues, and was used here to extract cues based on default features of the program, such as the part-of-speech tagger, and deception specific dictionaries added to the tool.

LIWC was used to extract the remaining cues. Linguistic Inquiry and Word Count (LIWC) (Pennebaker & Francis, 2001) processes text based on four main dimensions: standard linguistic dimensions, psychological processes, relativity, and personal concerns. Based on these dimensions, default dictionaries are available for 74 cues. Additional dictionaries can be added to either tool.

Table VI

Revised Constructs and Related Measures

Construct	Measurement
Quantity	Words, Verbs, Sentences
Specificity	Sensory ratio, Spatial ratio, Temporal ratio, Content Word Diversity, Bilogarithmic Type-Token-Ratio.
Uncertainty	Certainty Terms, Tentative Terms, Modal Verbs, Passive Voice, Generalizing Terms
Clarity	Redundancy, Sentence Length, Complexity Ratio, Average Word Length, Causation Terms.
Immediacy	1st person pronouns, 2nd person pronouns, 3rd person pronouns
Affect	Positive Activation, Negative Activation, Positive Imagery, Negative Imagery, Positive Pleasantness, Negative Pleasantness
Cognitive Processing	Exclusive Verbs, Motion Words, Cognitive Processing Terms.
Severity	Rating of incident severity in terms of related punishment

The Zhou/Burgoon framework included eight categories of cues. The revised framework included seven categories related to deception, in addition to the severity construct. The two frameworks are juxtaposed in Table VII below to highlight the differences between them.

Table VII

Summary of Zhou/Burgoon and Revised Frameworks

Zhou/Burgoon Framework		Revised Framework	
Construct	Related Cues	Construct	Related Cues
Quantity	Words, Verbs, Sentences	Quantity	Words, Verbs, Sentences
Specificity	1st person pronouns, 2nd person pronouns, 3rd person pronouns, other references, modifiers, sensory ratio and number of sensory details	Specificity	Sensory ratio, Spatial ratio, Temporal ratio, Content Word Diversity, Bilogarithmic Type-Token-Ratio.
Affect	Affect , Imagery, Pleasantness	Affect	Activation, Imagery, Pleasantness
Diversity	Lexical Diversity, Content word diversity, Redundancy	N/A	
Complexity	Average sentence Length, Average word length, pausality.	Clarity	Redundancy, Sentence Length, Complexity Ratio, Average Word Length, Causation Terms.
Uncertainty	Modal Verbs	Uncertainty	Certainty Terms, Tentative Terms, Modal Verbs, Passive Voice, Generalizing Terms
Nonimmediacy	Passive voice	Immediacy	1st person pronouns, 2nd person pronouns, 3rd person pronouns
Activation	Emotiveness, activation	N/A	
N/A		Cognitive Processing	Exclusive Verbs, Motion Words, Cognitive Processing Terms.
		Severity	Rating of incident severity in terms of related punishment

Confirmatory Factor Analysis

Confirmatory factor analysis was performed on the Zhou/Burgoon framework and the revised construct set. After constructing the models, absolute, incremental and parsimonious fit measures for the model were examined to assess the overall model. Absolute fit measures include the chi square statistic, goodness-of-fit index (GFI), and RMR, and RMSEA Absolute fit measures do consider any overfitting that may occur in the model. Incremental fit measures include TLI, NFI, and CFI. These measures assess the model fit compared to a null model that specifies no relation among the constructs or variables. The main parsimonious fit measure is the adjusted Goodness-of-fit index (AGFI), which assesses the fit of the model when considering the number of estimated parameters. The loadings of the constructs were examined for statistical significance. Each construct was also assessed for reliability. To assess construct reliability, alpha, composite reliability, and average variance explained were calculated (Hair, Anderson, Tatham, & Black, 1998).

Construct Validation Results

Traditional construct validation was applied to two competing models. The results below show that while a valid set of constructs was determined, some of the proposed cues and constructs may not be valid, particularly for this domain.

Model 1: Zhou/Burgoon Framework

As described previously, the Zhou/Burgoon model includes eight constructs with a total of 22 cues. The full model was first submitted to analysis; however, the solution was not admissible. The constructs were examined to determine which might be removed. Based on estimates of reliability, the complexity construct was removed, as calculations of alpha showed that there was negative covariance that could not be resolved. After removing the complexity construct, an admissible solution was generated. In that solution, all loadings were less than 1 and only one cue was not significantly loading, the 1st person singular pronouns. This cue was then removed. The resulting model was admissible, though the fit was not acceptable on all measures examined.

Table VIII

Zhou/Burgoon Framework Fit Measures

Fit Measure	Value	Suggested Value
RMSEA	0.10	<0.08
Adjusted Chi-Square	4.93	<5
Standardized RMR	0.08	<0.10
GFI	0.86	>0.9
CFI	0.90	>0.9
NFI	0.88	>0.9

The model resulting from performing CFA on the Zhou/Burgoon framework retained seven constructs and 17 cues. Four of the constructs—quantity, affect, diversity, and activation have acceptable construct reliability according to their composite reliability, average variance explained and alpha. Suggested values of these measures are 0.70 for composite reliability, 0.50 for average variance explained, and 0.70 for alpha. Specificity has poor reliability. This may be due in part to the large amount of zero values for 2nd person pronouns and 3rd person pronouns. This issue is discussed further in a later section. Reliability could not be assessed for Uncertainty and Nonimmediacy since these two constructs had only a single indicator.

Table IX

Zhou/Burgoon Constructs and Indicators Retained by CFA

Construct	Indicator	Standardized Loading	Composite Reliability	Average Variance Explained	alpha
Quantity	Word Quantity	1	0.94	0.83	0.92
	Verb Quantity	0.87			
	Sentence Quantity	0.86			
Specificity	1st Person Plural	0.53	0.42	0.17	0.33
	2nd person pronouns	0.36			
	3rd Person pronouns	0.5			
	Sensory Ratio	0.15			
Affect	Affect	0.12	0.81	0.66	0.74
	Imagery	0.98			
	Pleasantness	1			
Diversity	Lexical Diversity	1	0.80	0.62	0.74
	Content Word Diversity	0.87			
	Redundancy	0.3			
Uncertainty	Modal Verbs	1	1	1	-
Nonimmediacy	Passive Voice	1	1	1	-
Activation	Emotiveness	0.5	0.74	0.61	0.58
	Activation	0.98			

Model 2: Revised Framework

The second model analyzed was the revised framework. Similar to the Zhou/Burgoon framework, an admissible solution could not be found with confirmatory factor analysis when all variables and constructs were included in the model. Constructs and cues were removed one by one until an admissible solution could be calculated. The initial admissible solution included all proposed information except for the immediacy construct and the motion variable used as an indicator of cognitive processing. In this model, content word diversity and redundancy had loadings greater than one. Six additional cues did not have significant loadings.

Cues with loadings greater than one and those without significant loadings were removed, then replaced one at a time until these issues were resolved. In addition to the immediacy cues and the motion cue, the clarity and cognitive processing constructs had to be removed as no acceptable combination of cues could be found to represent these constructs. Additionally, the space, content word diversity, modal verbs, and passive voice cues had to be removed. The fit measures for the resulting set of constructs are shown below in Table X. As can be seen in the table, the values are acceptable for all of the fit measures.

Table X

Revised Framework Fit Measures

Fit Measure	Value	Suggested Value
RMSEA	0.08	<0.08
Adjusted Chi-Square	3.45	<5
Standardized RMR	0.06	<0.10
GFI	0.93	>0.9
CFI	0.94	>0.9
NFI	0.92	>0.9

Though the fit measures suggest that the set of constructs was good, when construct reliability is assessed, this was not the case, as only two of the four remaining constructs have acceptable values for the three measures being used here to assess construct reliability. Additionally, it should be noted that only four of the seven suggested constructs and 12 of the proposed cues have been retained, as shown in Table XI below.

Table XI

Revised Framework Constructs and Indicators Retained by CFA

Construct	Indicator	Standardized Loading	Construct Reliability	Average Variance Explained	Alpha
Quantity	Word Quantity	1	0.97	0.87	0.92
	Verb Quantity	0.87			
	Sentence Quantity	0.86			
Specificity	Sensory Ratio	0.34	0.35	0.17	0.33
	Temporal Ratio	0.24			
	Bilogarithmic Type-Token Ratio	0.57			
Affect	Activation	0.99	0.99	0.98	0.99
	Imagery	0.98			
	Pleasantness	1			
Uncertainty	Certainty Terms	0.53	0.54	0.30	0.46
	Generalizing Terms	0.74			
	Tentative Term s	0.29			

Construct Validation Summary

Confirmatory factor analysis has been performed on two sets of constructs on a data set including 366 statements. The results of this analysis are summarized in Table XVI. In both models tested, the Zhou/Burgoon framework and the Revised framework, the number of constructs and cues confirmed were less than what had been proposed. Two constructs, quantity and affect, were validated across both models, though the cues of these constructs varied somewhat. As affect is the only construct based on a previously validated set of constructs (Whissell, 1989), it is not surprising that it emerged as one of the constructs with acceptable validity. Quantity uses straightforward counts of words or sentences without relying on the definition or meaning of various words. It appears that the constructs based on lists or dictionaries of words are more problematic.

Table XII

CFA Result Summary

Item	Zhou/Burgoon framework	Revised Framework
Number of Constructs	7	4
Number of Reliable Constructs	4	2
Number of Cues	18	12
RMSEA	0.10	0.08
Adjusted Chi-Square	4.93	3.45
Standardized RMR	0.08	0.06
GFI	0.86	0.93
CFI	0.90	0.94
NFI	0.88	0.92

The Zhou/Burgoon framework retains more constructs and cues upon validation, though it does not have acceptable fit on most of the evaluated measures. As noted previously, three of the seven constructs retain only one or two indicators, preventing Alpha from being calculated for these three constructs. The revised model is the most parsimonious and has an acceptable fit, though both models assessed did not have acceptable reliability levels on all constructs. As model 2, the Revised model, is theoretically based, has a good fit, and is also parsimonious; it appears to be the best model. Using these results as a foundation, a set of hypotheses was developed to test the cues related to the constructs of the revised model. This is outlined in the next chapter.

CHAPTER IV

HYPOTHESIS TESTING

Hypotheses

One goal of this study was to identify the cues that distinguish truthful and deceptive messages in the high-stakes, real-world context. To accomplish this, hypotheses were developed for the cues representing each construct. However, the constructs of the study had not previously been validated. To ensure that hypotheses were made about appropriate constructs, factor analysis was conducted prior to hypotheses development. (See Chapter III for details of measurement of the constructs and construct validation). In addition to developing hypotheses regarding the difference between truthful and deceptive message groups on each linguistic-based cue, an additional hypothesis was developed regarding the impact of severity on these cues. The constructs presented below are those that could be validated in the revised construct framework. The remaining three constructs and the fifteen cues associated with these constructs are not presented in the hypotheses. The hypotheses presented here are summarized in Table XIII.

Quantity

Quantity reflects the length of the statement. According to IMT (McCornack, 1992), deceptive messages are edited such that sensitive information is omitted. This suggests that deceptive messages should demonstrate reduced quantity. Knapp (1974) also suggested that deceivers would exhibit reticence, including using fewer words, and confirmed this in their experiment. The self-presentational perspective concurs that deceptive statements will be shorter (DePaulo et al., 2003). Previous results regarding the construct of quantity have been mixed, with some suggesting deception increases quantity and others finding decreases. Therefore, it appears that this particular construct may apply differently depending on the domain or context.

Though these differing results may seem contradictory, it is actually consistent with IDT. This theory suggests that language will be adapted according to the context and the goals of the person producing the message (Burgoon et al., 2003). According to IDT, if time is available or efforts at persuasion may be beneficial, the deceiver is likely to create a longer message (Burgoon et al., 2003). This echoes the earlier finding of Watson and Ragsdale (Watson & Ragsdale, 1981). If the context is more conversational or the speaker provides a complete answer without interruption, he or she may increase quantity to appear believable or provide additional evidence to support his or her deception (Hancock et al., 2005). Similarly, if the deceiver is instructed to or is attempting to give enough evidence or detail to persuade the deceiver, it is expected that deceptive statements will show increased quantity as compared to truthful statements (Zhou, Twitchell et al., 2003).

The current context is believed to be closer to that where a person completes an answer without interruption and may increase quantity in order to appear believable. In one study where subjects were asked to discuss a topic completely, deceptive messages were shown to use more words than truthful messages (Hancock, Curry et al., 2004;

Hancock et al., 2005). Other studies have also supported the finding of greater word quantity in deceptive messages, including a text-based study where senders were motivated to convince the receiver that they were being truthful (Zhou, Burgoon, Nunamaker et al., 2004; Zhou, Twitchell et al., 2003). Based on this, the following is hypothesized:

Hypothesis 1: Deceptive statements will show greater quantity than truthful statements demonstrated by a)greater word quantity, b) greater verb quantity and c) greater sentence quantity.

Specificity

Specificity refers to language that establishes the context of the statement and perceptual information given (Zhou, Burgoon, Nunamaker et al., 2004). It describes the amount and type of details in a message. Reality monitoring posits that truthful messages will include more perceptual information since a subject is describing an actual experience (Bond & Lee, 2005; Zhou, Burgoon, Nunamaker et al., 2004). Contextual information includes language related to sensations experienced, spatial information and temporal information. This is consistent with the view of the self-presentational perspective which suggests that deceivers will provide less detail in their responses (DePaulo et al, 2003). Similarly, IDT proposes that deceptive statements will show reduced specificity. This may be expressed by reduced contextual detail and also by reduced lexical diversity, a measure similar to type-token ratio, reflecting less detailed content that truthful statements (Zhou, Burgoon, Nunamaker et al., 2004). The majority of studies in one analysis supported truthful statements having more spatial temporal and perceptual information than deceptive statements (Zhou, Burgoon, Nunamaker et al., 2004).

Bond and Lee also confirmed, using default LIWC dictionaries that deceivers tended to use less sensory information and less temporal information (Bond & Lee, 2005). Others have also shown deceivers to be less specific when forming their messages (Watson & Ragsdale, 1981). It was suggested that deceivers have trouble being specific about events or situations that do not exist. Experimental studies have confirmed that deceivers may be less specific by using fewer unique words in their messages (Knapp et al., 1974; Zhou, Burgoon, Nunamaker et al., 2004). This supports the following hypothesis:

Hypothesis 2: Deceptive statements will show less specificity than truthful statements, demonstrated by: a) lower sensory ratios, b) lower temporal ratios, and c) lower bilogarithmic type-token ratios.

Affect

For decades, it has been thought that deceivers would show greater negative affect. Newman and colleagues (2003) studied deception across five different samples. Consistently, deceivers used more negative emotion terms. This is in alignment with previous work that suggests that greater emotion may be present reflecting guilt or fear associated with lying (Ekman, 1985; Newman et al., 2003). These emotions may result in direct incorporation of affect in language, particularly negative affect (Zuckerman & Driver, 1987). Another study showed that deceivers used greater positive affect (Zhou, Burgoon, Nunamaker et al., 2004). Affect, as measured here, consists of three components: activation, imagery, and pleasantness, each of which are anticipated to increase in deceptive statements.

Hypothesis 3: Deceptive statements will show greater affect than truthful statements demonstrated by a) greater activation b) greater imagery, and c) greater pleasantness.

Uncertainty

Uncertainty may reflect attempts to avoid giving relevant answers. It includes the level of uncertainty, or strength and firmness of the passage (DePaulo et al., 2003; Fuller, Biros, Twitchell et al., 2006; Knapp et al., 1974). Knapp and colleagues were some of the earliest researchers to address linguistic cues. They proposed that deceivers would be more uncertain and also associated tentative constructions with uncertainty in deception (Knapp et al., 1974). IMT suggests that when the maxim of relation is violated, deceptive messages will fail to provide direct information. The messages will not include contextually relevant material that is expected. This information will be general, and will reflect the inhibited state of producing a deceptive message.

The self-presentational perspective also describes deceptive communications as being more uncertain. IDT suggests that deceptive messages will express uncertainty through more generalizing terms in an effort to deceive by ambiguity and evasiveness (Zhou, Burgoon, Nunamaker et al., 2004). The four-factor theory also predicted more use of generalizing terms. Previous results confirmed this (Buller & Burgoon, 1996; Knapp et al., 1974; Zuckerman & Driver, 1987), leading to the following hypotheses:

Hypothesis 4: Deceptive statements will show greater uncertainty than truthful statements demonstrated by: a) More generalizing terms, b) fewer certainty terms, and c) more tentative terms.

Severity

This study was conducted in a high-stakes context, or one in which there are consequences associated with being found either truthful or deceptive. Within the high-stakes context, there will be varying levels of consequences. Severity is used here to represent the consequences associated with a particular situation. It is a logical assumption that those who find themselves in severe situations will be motivated more than those in less severe situations. This should apply to both truthful and deceptive messages, as in both cases the sender of the message faces pressure to be believable. The self-presentational perspective of deception predicts that cues will be stronger when subjects are motivated to prevent lie detection. This perspective also predicts that lies about transgressions will result in stronger cues (DePaulo et al., 2003)

An analysis of previous deception studies (DePaulo et al., 2003) found that the cues to deception were clearer and more cues were significant when the deception was about a transgression; that is when deception was about a crime, mock crime, or similar situation. Zuckerman et al. found that language-related cues were particularly useful in detecting deception in motivated situations (Zuckerman & Driver, 1987). Zuckerman et al found that a greater number of cues were significantly associated with deception in a high motivation level as compared to low motivation. There were also a number of cues that were significantly different between high and low level conditions. Researchers have proposed that in high motivation situations, greater cue leakage will occur (Friedman & Tucker, 1990; Porter & Yuille, 1995). Previously, severity and its correlates have been measured as dichotomous. Here, it will be measured as a continuous variable. As the cues increase in strength when speakers are motivated and when transgressions are discussed, we propose that severity will have a positive relationship with the intensity of each cue:

Hypothesis 5: There will be a positive relationship between severity and cue intensity.

Table XIII

Hypothesis Summary

Hypothesis	Description
Quantity	
H1A	Deceptive statements will show greater word count than truthful statements
H1B	Deceptive statements will show greater verb count than truthful statements
H1C	Deceptive statements will show greater sentence count than truthful statements
Specificity	
H2A	Deceptive statements will show lower sensory ratios than truthful statements
H2B	Deceptive statements will show lower temporal ratios than truthful statements
H2C	Deceptive statements will show lower bilogarithmic type-token ratios than truthful statements
Affect	
H3A	Deceptive statements will show greater activation than truthful statements
H3B	Deceptive statements will show greater imagery than truthful statements
H3C	Deceptive statements will show greater pleasantness than truthful statements
Uncertainty	
H4A	Deceptive statements will show fewer certainty terms than truthful statements
H4B	Deceptive statements will show greater generalizing terms than truthful statements
H4C	Deceptive statements will show greater tentative terms than truthful statements
Severity H5	There will be a positive relationship between severity and cue levels.

Hypothesis Testing Methodology

Based on the validated constructs, and the cues that represent the constructs, MANOVA and Regression were used to test the hypotheses related to the deception constructs, as well as severity.

MANOVA

The second major goal of this study was to determine which verbal cues distinguish truthful and deceptive messages. For this component of the study, MANOVA was utilized. MANOVA is an appropriate statistical procedure to employ for the analysis of categorical independent variables and metric dependent variables that are at least interval scaled. It allowed us to analyze how the cues belonging to each construct separate truthful and deceptive subjects as a set and as individual cues (Hair et al., 1998). The maximum number of cues used to measure a single construct is three cues and the study included two groups, deceptive and truthful messages. The minimum recommended sample size for MANOVA is at least 20 observations per group or more observations per group than there are dependent variables. It has also been noted that achieving adequate power can be difficult with group sizes less than 50, however, this minimum will be exceeded for this study (Hair et al., 1998). The cell sizes were unequal due to the many more truthful that deceptive statements received. However, the software package used, SPSS, automatically adjusts for unequal cell size.

First, the model was assessed at the multivariate level to see if there was a significant difference of the vectors of means of the dependent variables across groups. This step was followed by the F-test to assess univariate differences for the individual cues. When using MANOVA the following assumptions must be checked: constant variance-covariance matrices occur across groups, multivariate normality of the dependent variables occurs within each group, observations are independent of each other, and the dependent variables are correlated.

Regression

To complement the MANOVA analysis used to study the difference between truthful and deceptive statements for each cue, the impact of severity on cue intensity was evaluated using linear regression. Based on the results of previous studies, it could be expected that more cues would be significant in situations of higher severity, and that there would be a significant difference between high and low severity conditions (Zuckerman & Driver, 1987). However, this implies a binary measurement of severity- either high or low. For this study, each statement received a severity rating between one and five. This scale, like many in the social sciences may be somewhere between ordinal and interval. The loss of information if treated as ordinal must be balanced against the error that may result if the scale is considered interval. In this case, the scale was treated as interval in order to perform more powerful statistical analysis (Pedhazur & Schmelkin, 2006).

A linear regression model was built for each of the twelve cues representing the four validated constructs. In each of the models, one of the twelve cues was the dependent variable. Severity and veracity condition, dummy coded for truthful or deceptive, were the independent variables. This design allows interpretation of the relationship between severity and the cue, as well as whether this impact is the same or different for truthful and deceptive statements.

Hypothesis Testing Results

Results for Between Group Differences

These results show the difference in cues between deceptive and truthful groups, testing the hypotheses described in Table XIII. So that assumptions of normality and homogeneity of variance could be met, several variables had to be transformed before hypothesis testing could be completed. The transformations used are listed along with their respective cues in Table XIV below. The multivariate test was significant for the quantity, specificity, and uncertainty constructs. This shows a significant difference in the vector of the means for these constructs, but does not show which group of statements, deceptive or truthful, has higher levels of any particular construct. To assess these differences, univariate tests were performed for each cue. The results show that the quantity related cues are significantly higher in deceptive than truthful statements, confirming hypotheses 1A, 1B, and 1C. Similarly, hypothesis 2B, and 2C were supported, as deceptive statements showed lower temporal ratio and lower bilogarithmic type-token ratio. Though there was a significant difference between the two groups in sensory ratio, the direction of the means was opposite of what was hypothesized. There were no significant differences for affect or its related cues. Therefore, hypotheses 3A, 3B, and 3C were not supported. While hypothesis 4A, and 4C were not supported, there was a significant difference between deceptive and truthful statements in the number of generalizing terms used, supporting hypothesis 4B. These results of the hypothesis testing are summarized in table XVI.

Table XIV

Summary of MANOVA Results

Cue (Transformation, if applicable)	Raw Mean (Std Dev)		Normed Mean (Std Dev)		Transformed Values Mean (Std Dev)	
	Truth	Deceptive	Truth	Deceptive	Truth	Deceptive
Quantity						
Word Quantity (Square Root)	**88.28 (84.70)**	**176.09 (102.19)**	**.16 (.18)**	**.34 (.21)**	**.35 (.18)**	**.56 (.18)**
Verb Quantity (Square Root)	**13.82 (16.82)**	**30.03 (22.18)**	**.13 (.16)**	**.29 (.22)**	**.32 (.18)**	**.50 (.20)**
Sentence Quantity (Square Root)	**5.56 (5.20)**	**9.99 (6.02)**	**.12 (.14)**	**.24 (.16)**	**.30 (.17)**	**.46 (.15)**
Specificity						
Sensory Ratio	2.94 (1.99)	3.51 (1.90)	.17 (.17)	.30 (.16)		
Temporal Ratio (Square Root)	**5.92 (3.82)**	**4.37 (2.38)**	**.27 (.18)**	**.20 (.11)**	**.48 (.19)**	**.43 (.12)**
Bilogarithmic Type-Token Ratio	**.90 (.03)**	**.89 (.02)**	**.46 (.17)**	**.41 (.12)**		
Affect						
Activation	1.39 (.61)	1.49 (.55)	.67 (.29)	.72 (.26)		
Imagery	1.30 (.58)	1.37 (.52)	.65 (.29)	.68 (.26)		
Pleasantness	1.51 (.66)	1.59 (.59)	.65 (.29)	.68 (.25)		
Uncertainty						
Certainty Terms (Square Root)	.39 (.91)	.46 (.76)	.95 (.12)	.94 (.10)	.97 (.08)	.97 (.05)
Generalizing Terms	**.93 (1.43)**	**1.58 (1.23)**	**.10 (.16)**	**.17 (.14)**		
Tentative Terms (Square)	1.38 (2.13)	1.42 (1.30)	.06 (.10)	.07 (.06)	.01 (.06)	.01 (.01)

*Bold print indicates significant result with Alpha=0.05

Results for Impact of Severity

Linear regression was performed to determine the impact of severity on cue intensity, or the amount of cue present in the statement (see Table XV for results). It was hypothesized that cues would intensify when severity increased. This was true for five of the twelve cues examined. For a sixth cue, bilogarithmic type token ratio, there was a significant relationship between severity and cue intensity, but in this case severity actually reduced cue intensity. Though not all results were significant, the results show that for most cues, severity tends to increase cue intensity. As shown by the deceptive coefficient, where truthful statements were dummy coded as 1, severity tends to increase more for deceptive that truthful statements, though not significantly so in 11 of 12 cases. For sensory ratio, severity significantly increases cue intensity, and does so significantly more for deceptive than truthful statements.

This portion of the study tested hypothesized differences between truthful and deceptive statements and the impact of severity on cue intensity. Seven of the hypotheses were supported, and an additional hypothesis was partially supported. Below, these results are compared to those of previous work.

Table XV

Summary of Regression Results

Construct	Cue (Transformation, if applicable)	Severity Coefficient	Deceptive Coefficient	R Square
Quantity	Word Quantity (Square Root)	**.40**	-0.05	.37
	Sentence Quantity (Square Root)	**.32**	-.03	.29
	Verb Quantity	**.35**	-0.04	.30
Specificity	Bilogarithmic Type-Token Ratio	**-.11**	.01	.04
	Sensory Ratio	**.16**	**-.07**	.14
	Temporal Ratio (Square Root)	-.07	.03	.02
Affect	Activation	.08	-.02	.01
	Imagery	.07	-.01	.01
	Pleasantness	.08	-.00	.01
Uncertainty	Certainty Terms (Square Root)	-.03	-.01	.01
	Generalizing Terms	**.10**	-.03	.06
	Tentative Terms (Square)	-.01	.00	.00

Summary of Between Group Differences

Significant differences in the expected direction were found for seven of the twelve cues examined. This study found that deceivers will use greater quantity, indicated by greater word, verb and sentence quantity. This is consistent with the findings of at least one previous study (Zhou, Burgoon, Nunamaker et al., 2004) and the prediction that deceptive statements would show increased quantity due to attempts to provide convincing evidence. Two studies using the desert survival task (Zhou, Burgoon, Nunamaker et al., 2004; Zhou, Twitchell et al., 2003) found that deceivers use significantly less diversity in language, which is consistent with the finding of lower bilogarithmic type-token ratio. Kohnken (1985) also found lower type-token ratio in deceptive speech of eyewitnesses. It appears that quantity-related cues and bilogarithmic type-token ratio are appearing in a manner similar to at least some other works, as predicted. The lack of significance for most cues related to affect and uncertainty is also consistent. So while there may be some aspects of deception that vary with domain, it appears that some cues are consistent.

Some previous studies (Zhou, Burgoon, Nunamaker et al., 2004; Zhou, Burgoon, Twitchell et al., 2004) have separated affect into positive and negative dimensions. Though construct validation strongly supported affect as measured here, measuring positive and negative affect separately may be worth exploring, since significant differences were not found here. Similar to the current study, a previous experiment failed to find uncertainty terms to be significantly different between the groups (Sporer, 1997). If this continues to be the case, alternative or larger dictionaries may be needed to measure the construct of uncertainty. As described in further detail in a later section, there was a large number of statements for which certainty terms were not present, which certainly could impact any ability to find significant results.

Only one cue, sensory terms, had a significant difference in a direction opposite of what was predicted. There is no clear explanation for this; however, it is possible that this cue corresponds somewhat with quantity. That is, subjects may inadvertently include more sensory information, which may or may not be accurate, in an effort to sound convincing.

Table XVI

Summary of Hypotheses Results

Hypothesis #	Hypothesis	Result
Quantity		
H1A	Deceptive statements will show greater word count than truthful statements	Supported
H1B	Deceptive statements will show greater verb count than truthful statements	Supported
H1C	Deceptive statements will show greater sentence count than truthful statements	Supported
Specificity		
H2A	Deceptive statements will show lower sensory ratios than truthful statements	Not Supported
H2B	Deceptive statements will show lower temporal ratios than truthful statements	Supported
H2C	Deceptive statements will show lower bilogarithmic type-token ratios than truthful statements	Supported
Affect		
H3A	Deceptive statements will show greater activation than truthful statements	Not Supported
H3B	Deceptive statements will show greater imagery than truthful statements	Not Supported
H3C	Deceptive statements will show greater pleasantness than truthful statements	Not Supported
Uncertainty		
H4A	Deceptive statements will show fewer certainty terms than truthful statements	Not Supported
H4B	Deceptive statements will show greater generalizing terms than truthful statements	Supported
H4C	Deceptive statements will show greater tentative terms than truthful statements	Not Supported
Severity		
H5	There will be a positive relationship between severity and cue intensity.	Partially Supported

Summary of Severity Regression Analysis

There are two main studies that have previously examined concepts similar to severity in deception research (DePaulo et al., 2003; Zuckerman et al, 1987). Depaulo and colleagues (2003) studied the impact of trangressions, such as crimes and other misdeeds, on cues and found that the cues were clearer for 11 of the 12 cues examined. That study found that response length had greater magnitude for transgressions versus lies not about transgressions. This was the only cue examined that could be a linguistic-based cue. Zuckerman et al. (1987) studied the difference in high and low motivation conditions for a number of deception cues, including five verbal cues: negative statements, irrelevant information, self-references, immediacy, and leveling terms. In the low motivation condition, deceivers used significantly fewer immediacy terms. In the high motivation condition, deceivers used significantly more negative statements and levelers.

There were no cues comparable to negative statements and immediacy terms included in the hypothesis testing. Levelers are conceptually similar to generalizing terms, used in this study as an indicator of uncertainty. Zuckerman's finding of more levelers in deceptive behavior is consistent with the finding of more generalizing terms in deceptive statements in this study. Though not all results were significant, these exploratory results show that severity can have an impact on cue intensity for both truthful and deceptive statements in a high-stakes environment.

This chapter presented a set of hypotheses related to the validated constructs. Specifically, hypotheses were developed for the twelve cues related to the constructs of quantity, specificity, affect, and uncertainty. A hypothesis was also developed to describe the relationship between severity and cue intensity. The methodology and results for the hypothesis testing were also described. The next chapter will present the methodology, results and analysis for the final piece of the dissertation, the classification models used to determine veracity.

CHAPTER V

CLASSIFICATION METHODOLOGY, RESULTS, & ANALYSIS

Methodology

Message Feature Mining

The final goal of this study was building a decision support system to identify deceptive messages using linguistic-based cues. Essentially, this step included building a variety of models for the purpose of classifying truthful and deceptive statements, following a process known as Message Feature Mining (Adkins, Twitchell, Burgoon, & Nunamaker Jr, 2004), outlined in Table X. This process has two main steps: extracting features and classification. Key aspects of the feature extraction phase were choosing appropriate features, or cues, and calculating those features over desired text portions. This entailed processing the text through appropriate programs in order to quantify the levels of the linguistic cues present in the statements.

After completing this step, five of the 371 statements were excluded from the sample due to excessively short or long length, leaving 366 statements to be used for classification. Key components of the classification phase are choosing an appropriate classifier, and training and testing the model. Logistic regression, decision trees, and artificial neural networks were selected as the classification methods to be used due to their common use and their use in a previous study in automated text based deception detection (Zhou, Burgoon, Twitchell et al., 2004)

Table XVII

Message Feature Mining Process

Main Steps of Message Feature Mining Process
1. Select desired features or cues
2. Identify and quantify features in text using text processing tools
3. Select types of classification models to be built
4. Train and test models
5. Evaluate model performance
6. Identify important features

Several cue sets were used to develop alternate classification models in order to identify the best set of inputs and the most accurate model. The sample was balanced in order to obtain better performance with the various data mining algorithms (Berry & Linoff, 2004). Here, the number of deceptive statements was the limiting factor in balancing the data set. There were 79 deceptive statements and 287 truthful statements. While software may automatically balance the data, if the reduced sample is not carefully constructed, bias may be introduced into the results. To overcome this, the 287 truthful statements were randomly divided into four partitions containing 71 to 72 statements. Four data sets were then formed, each including all of the deceptive statements and one of the four partitions of truthful data. Three data sets included 151 statements and the fourth had 150 statements. Classification models were constructed using each of the four data sets.

Cue Sets

The size of the data set was an important consideration in choosing a subset of the more than 30 cues available to be used for building the classification models. Based on artificial neural network heuristics (Sarle, 2004), if the entire set of available cues were used, well over 600 statements would be needed to achieve generalizable results. Of the 120 samples reported by DePaulo et al (2003), the largest sample size was 192. Further, due the difficulty of establishing "ground truth", it is not likely that a sample of 600 or more items could feasibly be collected. For these reasons, investigating methods for limiting the number of variables into classification models for automated deception detection has additional merit.

While other types of classifiers are not as restrictive as the neural network in terms of number of inputs, the same data and cues were presented to each model so results would be comparable. A previous study using linguistic based cues included 22 variables extracted using A99A as classification inputs (Zhou, Burgoon, Twitchell et al., 2004). This represents the largest cue set used for building classification models for deception detection.

The results of this previous study showed that accuracy was improved when the original set of 22 variables was reduced to only those variables identified as important after training and testing the models on the full set of 22 cues. To reduce the set of 22 cues, variables that were listed as important in classifying messages for at least 2 of 4 techniques were identified. This provided a list of 14 cues, listed in Table XVIII, which form the first of the four cue sets, which will be referred to as the Zhou/Burgoon cues. The Zhou/Burgoon cues showed reasonable accuracy in a previous study and will provide a comparison point for the current work. However, that study was a laboratory study in which students discussed the well-known desert survival problem. Since the domain was changed for this study, those cues may not translate to a new situation.

The need for theoretically based deception detection methods drove the selection of the second set of variables (Council, 2003). A set of deception constructs for use in linguistic analysis were identified and validated in Chapter III. The twelve cues that serve as indicators of the four constructs compose the second cue set. The use of this second, construct related, cue set may determine the suitability of cues selected strictly for their theoretical origin, not for their applicability for data mining analysis or previous use in classification studies.

It is likely that an accurate model could be achieved with one of the first two cue sets. However, in order to determine the best set of cues, a third set, the comprehensive cue set was implemented to identify the best combination of cues from all that were available. For the third set of cues, a list of cues including the cues identified in the A99A studies, the validated framework cues, and previous studies implementing LIWC to study deception was compiled. This included 31 cues, which have been labeled the comprehensive cue set.

A feature selection procedure was used to develop the fourth cue set. The feature selection was applied to the comprehensive cue set to reduce this overall list to a number more appropriate to the size of the data set. The specific procedure used the f-statistic to determine the relationship between a given cue and the dependent variable. The variables were then ranked in importance according to this relationship. The eight most important cues were retained to form the fourth and final cue set. Based on the size of the data set and neural net heuristics (Sarle, 2004), it was determined that this cue set should be limited to eight variables in an attempt to maximize generalizability of results.

There are of course several other feature selection procedures that could be implemented, but using the f-statistic for feature selection was deemed to be a reasonable starting point due to its simplicity and availability. The classification results based on this last cue set can be used to explored the utility of additional feature selection methods. The four cue sets are summarized in Table XVIII.

Table XVIII

Summary of Classification Cue Sets

Cues	Zhou/ Burgoon Important Cues	Comprehensive Cue Set	Text-Based Deception Construct Cues	Feature Selection Cues
1st person plural pronouns	X	X		
1st person singular pronouns	X	X		
2nd person pronouns	X	X		
3rd person pronouns		X		X
Activation	X	X	X	
Average sentence length		X		
Average word length	X	X		
Bilogarithmic type-token		X	X	
Causation terms		X		
Certainty terms		X	X	
Cognitive processing		X		
Content word diversity	X	X		X
Emotiveness		X		
Exclusive terms		X		X
Generalizing terms		X	X	
Imagery	X	X	X	
Lexical diversity		X		X
Modal verbs		X		
Modifiers	X	X		X
Motion terms		X		
Passive verbs		X		
Pausality	X	X		
Pleasantness	X	X	X	
Redundancy		X		
Sensory ratio	X	X	X	
Sentence quantity		X	X	X
Spatial ratio	X	X		
Temporal ratio	X	X	X	
Tentative terms		X	X	
Verb quantity	X	X	X	X
Word quantity		X	X	X

Classification Models

Following cue selection and text processing, the next main step of Message Feature Mining was choosing the appropriate classification method. This was followed by training and testing the models, then evaluating the model performance. The common classification methods used were artificial neural network, decision tree, and logistic regression.

An artificial neural network is a system of connected units, or nodes, which are arranged in layers. Typically, an artificial neural network has an input layer, a hidden layer, and an output layer. The nodes in the hidden layer combine the inputs from the previous layer into a single output value which is passed on to the next layer. Associated with each unit in the network is a weight. The weights in the network are determined by training the network on a portion of the data. The network performance is then tested on the remaining data, or holdout sample (Berry & Linoff, 2004). The network that was utilized was the common feedforward multilayer perceptron.

Though artificial neural networks have been shown to be powerful classifiers whose performance may exceed other classifiers, in terms of accuracy, artificial neural networks are widely considered to be ‘black boxes’ that do not readily give an explanation as to precisely how the classification decisions are made. For artificial neural networks, sensitivity analysis (Engelbrecht, Cloete, & Zurada, 1995) may be used to calculate variable importance. Similar procedures have been applied to automated deception detection using linguistic analysis in the past (Zhou, Burgoon, Twitchell et al., 2004).

The second classification method used was a decision tree algorithm. Decision trees function by dividing a set of records into successively smaller sets by applying a set of decision rules. There are a variety of methods that can be used to determine the best way to split the record set, such as entropy reduction, gini, information gain, or the chi-square test. This study used the maximum information gain criteria. Decision trees provide a user-friendly set of if-then rules that can be used to classify data (Berry & Linoff, 2004). C5.0 was the specific decision tree algorithm implemented in this study. This algorithm builds a tree, then prunes it to produce a more generalizable tree (Berry & Linoff, 2004). The result of the decision tree algorithm is both the classification of all data items as truthful and deceptive and a set of if-then rules that can be used to explain how these classifications were made.

Logistic regression is a statistical technique appropriate for use with continuous independent variables and a binary dependent variable. Although discriminant analysis could also be considered in these circumstances, logistic regression does not face the same strict assumptions, so it may be useful under a wider range of circumstances. In logistic regression, the Wald statistic can be used to assess significance of individual cues (Hair et al., 1998). Further, standardized coefficients give an indication of variable importance. SPSS Clementine was used for the classification portion of message feature mining. Though WEKA has been used previously as part of A99A for classification, Clementine was used here due to additional output details that are available.

To ensure that the results were not due to the particular train and test samples selected, ten-fold cross validation was implemented on each of the four data sets. The data set was first partitioned into ten equal sections. Nine sections were used for training the appropriate model and the remaining section was used for testing the model. This process was repeated ten times, so that each of the ten sections of the data set was used once as the testing set. The partitions were stratified so that the observations were split approximately equally between the two possible outputs, truthful and deceptive within each partition. Cross validation provides better estimations of the true error rate of the classification model than a single train-and-test experiment (Weiss & Kulikowski, 1991). The results of the ten experiments for the four data sets were aggregated to estimate the overall accuracy of each classification model.

Each of the classification models provides somewhat different information. For example, logistic information provides information on the significance of each cue and whether the cue has a positive or negative relationship with the dependent variable. Without advanced algorithms, only the relative importance of each cue in a neural network model can be extracted. Despite this, some results can be consistently pulled from each model, including overall accuracy, sensitivity, specificity and false positive rates. The focus of the analysis is these measures that can be compared across algorithms. These measures were compared for each pair formed by the three algorithms and four cue sets.

MANOVA was used to determine if there are significant differences due to cue set used, model, or the interaction of these two factors on the dependent variables-overall training data accuracy, overall test data accuracy, sensitivity, specificity, and false positive rates. Overall accuracy measures the overall percentage of cases correctly classified for either the training or testing data. Sensitivity measures the true positive rate, or percentage of actual deceptive cases correctly predicted. Specificity measures the true negative rate, or percentage of actual truthful cases correctly classified. Additionally, false positive levels were assessed. This is the ratio of actual truthful cases predicted as deceptive to the number of actual truthful cases. These measures can be assessed for artificial neural network, decision tree, and logistic regression models.

Cue Importance

For each type of classification model, different methods were used to determine which of the cues were the most important discriminators. In addition to analyzing the accuracy of the models, the importance of the cues used in each model was evaluated. For each model, the importance of each cue within individual models was evaluated. Additionally, the number of times a cue appeared as important in the ten iterations of cross validation was also considered. For the logistic regression and decision tree models, only important variables are included as the model is built. For these models, determining which variables are important is then relatively straightforward. For the neural network model, all variables are retained by the model by default, though sensitivity analysis can be performed to determine the relative importance of each cue. To determine which should be included in the list of important variables for a neural network, there is no clear cutoff. Here, the number of variables retained by the decision tree and logistic regression models were considered. The sensitivity analysis results were also evaluated to determine if a clear cutoff point emerged.

Overall Classification Results

The classification models described above were evaluated on several performance measures and the importance of the various cues for different model and cue set combinations were ascertained. MANOVA was used to determine if there were significant differences in the classification performance measures for any model, cue set, or model/cue combination. At the multivariate level, the overall model was significant. There was a significant main effect for cue set for sensitivity. Post hoc contrasts showed that the Construct cue set was better than the feature selection cue set, which was better than the Zhou/Burgoon dataset and the Comprehensive set. There was also a significant interaction between technique and cue set for training accuracy. For the remaining measures, there were no significant main or interaction effects.

Cue Set Results

Cue Set 1: Zhou/Burgoon cues

The first set of variables analyzed was the Zhou/Burgoon Cues. The levels or amounts of these 14 cues were extracted using GATE. For example, the number of verbs and average word length were output from the text processing program. They were then used to build the three types of classification models. As described previously, all results were calculated using ten-fold cross validation for each of the four data sets. To assess the performance of the classification model, overall classification accuracy for training and testing data, false positive rates, sensitivity and specificity were analyzed (See Table XIX for a summary of these results).

Using this set of cues, the neural network model has the greatest overall accuracy for the test data and the decision tree model has the lowest false positive rate. For the neural network model, there was a large difference in the training and testing data accuracies. All three models showed significant differences on the training data accuracy with thiscue set, with the neural network having the highest accuracy and logistic regression having the lowest accuracy.

Table XIX

Summary of Classification Results

Measure	Logistic Regression Mean (Std Dev.)	Model Decision Tree Mean (Std Dev.)	Neural Network Mean (Std. Dev)
Overall Accuracy % (Train)			
Zhou/burgoon	72.54 (4.78)	82.50 (5.52)	91.26 (4.12)
Constructs	75.50 (3.70)	81.25 (6.93)	87.42 (5.37)
Comprehensive	79.46 (6.06)	86.07 (4.48)	97.07 (5.62)
Feature Selection	72.04 (4.45)	77.69 (4.71)	84.60 (5.10)
Overall Accuracy % (Test)			
Zhou/burgoon	66.93 (11.89)	69.59 (10.68)	69.89 (9.13)
Constructs	71.16 (10.41)	71.19 (11.21)	73.86 (8.23)
Comprehensive	70.51 (12.66)	67.12 (11.09)	70.46 (12.77)
Feature Selection	69.34 (11.39)	70.87 (10.80)	72.01 (11.53)
False +			
Zhou/burgoon	30.76 (20.64)	29.29 (22.39)	31.30 (17.61)
Constructs	27.05 (15.32)	33.39 (18.11)	32.55 (15.67)
Comprehensive	30.54 (19.52)	38.57 (18.99)	29.33 (18.47)
Feature Selection	30.27 (18.19)	37.01 (17.02)	31.25 (18.89)
Sensitivity			
Zhou/burgoon	65.00 (14.15)	68.75 (16.60)	71.03 (15.49)
Constructs	69.46 (16.05)	75.40 (19.45)	79.69 (12.81)
Comprehensive	71.47 (14.34)	72.28 (12.88)	70.49 (19.05)
Feature Selection	68.93 (16.36)	77.95 (16.33)	74.87 (15.49)
Specificity			
Zhou/burgoon	69.24 (20.64)	70.71 (22.39)	68.71 (17.61)
Constructs	72.95 (15.32)	66.61 (18.11)	67.46 (15.67)
Comprehensive	69.46 (19.52)	61.42 (18.99)	70.67 (18.47)
Feature Selection	69.73 (18.19)	62.99 (17.02)	68.75 (18.89)

Cue Set 2: Text-Based Deception Constructs

Next, the set of 12 cues used as indicators of the four validated constructs were used as inputs to the classification models. The neural network model has the highest test accuracy rate (73.86 percent) for all models and cue sets. The logistic regression model shows the lowest false positive rate for all models and cue sets. The decision tree model has the highest false positive rates. For this set of cues, which includes two less cues than cue set one, the reduction in accuracy from training to test data sets is smaller than cue set one. Again, the three models had significantly different results on training accuracy.

<u>Cue Set 3: Comprehensive Cue Set</u>

For the third set of cues, the highest accuracy was found for the logistic regression model. The decision tree model has the most false positives and lowest overall test accuracy, suggesting that is the worst technique to apply to the comprehensive set of cues. Like the first two cue sets, each model was significantly different from the other two models for training accuracy.

<u>Cue Set 4: Feature Selection Cues</u>

The fourth set of classification models were built using the eight cues that were selected using the feature selection procedure. The neural network model is most accurate, and has an intermediate false positive rate. The decision tree model had significantly better training accuracy than the logistic regression model for this cue set. For the training data set, the accuracy of logistic regression and decision tree models were each significantly different from the neural network model, though there was no significant difference between the logistic regression and tree models.

Summary of Classification Results

The best result in this study using a large cue set was 73.86 percent for a neural network model, exceeding the maximum accuracy of 71.1 percent found in previous studies relying on LIWC alone for text processing. However, this accuracy rate of nearly 74% was not significantly better than the accuracy of any other cue and model set combination. At this point, this method is not as accurate as the polygraph. It is possible that if additional data could be added to the sample the accuracy could be increased.

Alternative classification models might also be used that could improve accuracy. However, with a maximum accuracy approaching 74 percent, it is well within the 72 to 92 percent accuracy shown in polygraph field studies (Council, 2003). For three of the four cue sets, the neural network model provides the highest test accuracy, while the logistic regression model has the lowest test accuracy for three out of four cue sets. The differences in test accuracy were not statistically significant for the models, cue sets, or the interaction between the two.

Though the cue and model set differences were largely not significant, the differences are practically significant. For example, there is a difference of about seven percent in test accuracy between the top model, the construct cues neural network model, and the worst model, the Zhou/Burgoon logistic regression model. To those determining which person in a group of individuals involved in a crime is telling the truth or being deceptive, that 7 percent accuracy difference might be quite important. While definitive conclusions cannot be made, the finding of highest accuracy for a neural network is similar to what was found in the Zhou/Burgoon dessert survival study and a small set of cues seem to be emerging as important across studies. This result can provide guidance, as well as a basis for comparison for future studies.

Summary of Cue Importance

In addition to analyzing the performance of the models, the importance of the cues used in each model was evaluated. Since our interest is in identifying the cues with the greatest capacity to distinguish truthful and deceptive statements, the important cues are only reported for the most accurate of the four data sets for each model and cue set combination. For each of the twelve model-cue combinations, the importance of each cue within individual models was evaluated. Additionally, the number of times a cue appeared as important in the ten iterations of cross validation was also considered. For the logistic regression and decision tree models, only important variables are included as the model is built. For these models, determining which variables are important is then relatively straightforward. For the neural network model, sensitivity analysis was used to determine which cues were the most important. The important cues are summarized in Table XX below.

Word quantity, verb quantity, and sensory ratio are the only cues that matter for all of the classification models. These three cues, along with temporal ratio, were important for at least three of the four cue sets. Since these cues appear to work well regardless of the model or other cues used, future studies may focus on these three variables. Thirteen additional cues are important for at least one model. As is shown in the table below, fifteen of the variables were not important in any of the models for any cue set. Cue set 3, the comprehensive set of cues, basically subsumes the other three cue sets. Therefore it is not surprising that it is the cue set that most often shows overlap with other cues sets regarding which cues are important for a given type of model.

Table XX

Summary of Cue Importance

Cue	NN	DT	LR
1st person plural pronouns	A2	Z2	
1st person singular pronouns		Z2	A4
3rd person pronouns		F2	
Activation		A2	
Bilogarithmic type-token ratio			A4
Emotiveness	A2		
Generalizing terms		A2, C2	
Imagery	Z4	C2	
Lexical diversity	A2		A4
Motion terms	A2		A4
Pleasantness		A2	
Sensory ratio	C1	C2	A4, C1, F1
Spatial ratio		Z2	
Temporal ratio	C1, A2, Z4		A4
Verb quantity	F4, Z4	Z2	A4, Z2
Word quantity	C1, F4	A2, C2, F2	F1, Z2, C1
Variables not important for any method: average sentence length, average word length, causation, certainty terms, cognitive processing terms, content word diversity, exclusive terms, modal verbs, modifiers, passive verbs, pausality, redundancy, second person pronouns, sentence quantity, tentative terms			
Table indicates cue set (A=All, C=Construct,F= Feature Selection, Z=Zhou/Burgoon)and partition(1,2,3,4)			

In comparison to the Zhou/Burgoon classification study (Zhou, Burgoon, Twitchell et al., 2004), most of the variables found to be important in that study were also relevant in this study. Only 2nd person pronouns, average word length, modifiers, and pausality were previously found important, but were not important here. For this particular sample, it is not surprising that second person pronouns, the ‘you’ pronouns, were not important. In this case, the ‘you’ would be law enforcement personnel. As the statement is generally about a previous incident in which law enforcement were not present, it would not be logical to use these pronouns.

There was some variation from that study regarding which technique the cues were important for. Several cues that were not included in the list of important variables from the original set of 22 variables in the Zhou/Burgoon study also failed to be among the most important in this study. This included: average sentence length, modal verbs, passive voice, and redundancy. For the Zhou/Burgoon desert survival study, approximately five to seven cues emerged as important for each model. Here, two to seven cues were important per model. As evidence accumulates across domains as to which variables are important, or not, this can aid researchers in narrowing down the large list of potential cues and determine how many cues to include.

CHAPTER VI

DISCUSSION AND CONCLUSION

Contribution to Literature and Practice

Traditional construct validation procedures, including confirmatory factor analysis (CFA), were used to validate a set of constructs for use in researching text-based high-stakes deception. These results are expected to generalize to other high-stakes domains. However, several proposed constructs could not be validated, including constructs previously used without proper validation. This shows the importance of validating constructs for each domain, as all constructs suggested by theory may not be widely applicable. The reason that several cues and constructs could not be confirmed may be due to measurement or domain. Regardless of the reason, this process must be repeated with additional samples to investigate this issue.

This study also showed differences between deceptive and truthful statements on several cues examined. Theoretical predictions were confirmed for several cues related to quantity and specificity. However, the findings also suggest that the current method for measuring some cues may need improvement. This study has also shown that severity is an important issue to investigate further. Here, severity has been studied in a simplistic, exploratory manner. Though the method was not sophisticated, it was successful, in that it was shown to significantly impact cue intensity for several cues. These findings show that severity, or high-stakes certainly is an important issue that should receive further attention and it is also relevant in studying text-based deception.

The classification results show that accurate automated-text based deception detection can be accomplished with high-stakes, real-world data. This is an important finding with implications for law enforcement, human resources personnel, and others. The results also show that classification can be done with parsimonious cue sets. The most important cues were identified, some of which were also important in previous studies. There were also a number of cues that were previously found to be important that did not play a large role here. The findings here suggest that while some cues will differ from domain to domain, there are a few that are emerging as important across domains, while others do not seem to enter the model regardless of the context. Given these results, this line of research should move forward, since this has the potential to fill the need for portable, user-friendly, unobtrusive deception detection in the field.

Limitations

As described above, some cues, particularly those related to pronoun usage appear to be problematic. An analysis of descriptive statistics for the sample showed that for several variables, a value of zero was recorded for a large number of statements. This may impact the overall analyses, as this will severely limit the number of non-zero data points for a given indicator. Clearly, if a construct is not present in the data, it cannot be validated. It should be noted that these zero values can be distinguished from missing data. Missing data are those values for which the value is unknown. Here, we know that the value is zero, indicating the type of language measured by a given indicator is simply not present in the current sample. Variables for which the data had more than 60% zero values are shown in Table XXI below.

Table XXI

Cues with Large Number of Zero Values

Variable	% of Records with value of 0
Causation Terms	65.58
Certainty Terms	72.90
Modal Verbs	73.44
Passive Verbs	73.44
1st Person Plural Pronouns	82.66
2nd Person Pronouns	92.14

The lack of certain types of pronoun usage is likely due to the domain. When writing a statement, subjects are expected to use the name of each person involved in the situation. Therefore, they are more likely to say ‘Bob and I’ in an instance that they might normally say ‘we’. Similarly, the lack of ‘you’ references or second person pronouns may be limited by the context. If a person were to use these pronouns, they would have to be referring to the person or people who required them to write the statement. As these people were most likely not involved in the incident as it took place, they would not be referred to in the description of the incident. There is not a clear reason for the limited use of passive and modal terms in the statements.

The frequency at which zero values appear has not been previously reported, so this may or may not be related to the domain. The use of causation terms has only been reported in one previous study (Hancock et al., 2005) and the certainty dictionary has not been previously used in deception research. It is possible that these are not valid cues for deception research or they may not be relevant in this domain. Additional research including alternate samples will be required to make this determination. The zero values also impacted the shapes of the variable distributions, impacting the MANOVA analysis. Several transformations were necessary. The certainty variable was one with many zero values. There was no significant finding for this variable, perhaps impacted by the zero values. Some of the variables with large numbers of zeros were included in the classification models, because either they were theoretically confirmed or to allow comparison with previous work. These variables included: certainty, 1st person plural pronouns and second person pronouns. The comprehensive cue set, of course, includes all of the cues. Aside from 1st person plural pronouns, the cues with large numbers of zero values were not important in any of the models.

Most deception research to date has been conducted in the laboratory, producing inconsistent results (DePaulo et al., 2003). Field research offers the opportunity for realism, but does lack the control afforded by the laboratory. Despite sacrificing key experimental elements such as randomization and manipulation, this is outweighed by the need for real-world data, as sufficiently and realistically replicating a high stakes environment may not be possible in the laboratory.

As has been the case in previous deception research, the available sample size may have limited the accuracy that could be achieved. As this is not something that is likely to change, this will heighten the importance of selecting valid and appropriate cues. A key factor limiting the sample size is the difficulty in determining ground truth. Here, every effort was made to correctly identify statements as truthful or deceptive, though this process cannot be full proof. In these situations it is better to err on the side of caution and exclude questionable statements, as was done here. While it is anticipated that this study will generalize to other high-stakes situations, additional studies will be necessary to verify this. As has been found in the past, it is somewhat unlikely that the results will generalize to all deceptive situations. Therefore, defining deception and the related cues for well-defined contexts is a more pertinent goal than defining a universal cue set and classification model.

Future Directions

This research has validated a set of deception constructs in a real world-high stakes domain. It is not expected that the results found here will generalize across all domains. These results may very well translate across different high stakes situations. Future studies should include alternate samples, preferably from the real world, that allow comparison of samples with varying stakes with the current sample. Future studies will expand upon the domain studied to explore the limits to the generalizability of this research and determine which constructs apply to a given context. Further validation of these deception constructs are only part of future research to be conducted. Studies will also be needed in order to determine whether the findings which failed to validate some constructs are due to the domain and noisiness of field data or if it is the underlying theory itself that is faulty. It may also be that some cues have not been sufficiently defined and additional dictionaries will need to be developed to more fully capture the related concepts.

Like this study, previous studies have analyzed the difference in cues between truthful and deceptive groups. A thorough comparison of the results from this and previous studies can show which cues operate the same or differently across domains. In addition to reporting means for truthful and deceptive statements, additional reporting of the number of data points with zero values can provide further insight. While high and low motivation states have been used in previous deception research, this study is one of the first to measure severity on a continuous scale, addressing the previously identified need for high-stakes deception research. The results here show that severity impacts cue intensity for six of the twelve of the cues studied. Given this initial success, severity should be measured in future studies to further understand how this concept impacts cue intensity. The measure of severity used here was quite simple. In a real-world setting, it may be quite difficult to measure severity in more complex ways, particularly when historical data is used. However, in a laboratory situation, a more complex measure might be possible. Different measures could be compared to assess whether a simple measure is sufficient or a more complex measure is indicated.

This study showed reasonable accuracy in the first relatively large-scale attempt to implement automated text-based deception detection using field data. However, it is believed that these results can be improved. No clear pattern has yet emerged to definitively suggest the best set of cues or the best type of classifier, though a small set of cues has remained important across two studies. MLP neural network models have also consistently performed well across these studies. However, since the performance of the MLP was not significantly better than other models, we cannot yet rule out the use of any of the traditional classifiers in this stream of research.

As the results suggested better performance with smaller cue sets, additional methods for choosing the best reduced cue set should be explored. The feature selection used here based on the f-statistic may complement logistic regression since it is a typical statistical method relying on linear patterns. Methods which recognize both linear and nonlinear relationships between the inputs and outputs of the model may increase accuracy of the decision tree and neural network models. Further, additional classifiers such as radial basis function neural networks, random forest, and boosted decision trees will be considered.

The portion of this study with the highest time requirement was transcribing the written statements. Alternative methods of capturing written information will be explored. This might include having suspects use tablet pc's or traditional word processing software to record their statements. As mentioned previously, it is expected that the relevant cues to deception will differ by domain. Samples from other domains and cultures can provide assistance into defining the cues to use with specific samples and also show if some cues are generalizable.

Conclusion

This study is the largest known study to examine text-based deception in a real-world domain. A set of constructs for this domain were validated and the notion of context-specificity in deception research was affirmed. For many of the cues examined, expected differences between truthful and deceptive statements were found. However, it was shown that when examining truthful and deceptive statements in a high-stakes situation, the impact of severity on cue intensity must be examined. This study also found that real-world deception data can be accurately classified using a combination of text-processing programs and data mining software. Additional classification methods and cue selection procedures need to be explored to increase classification accuracy to the point that the technique can be employed in the field. Additional field studies should be conducted to determine whether the results here are unique to the real-world domain or to the high-stakes context, or some combination of these factors.

This study focused on actual deception. Most studies have focused on either actual or perceived deception. Additional knowledge could be gained by studying both actual and perceived deception within a single high-stakes study. Regardless of form of analysis, this study shows that the findings from deception research must be compared across domains so that a more complete picture of deception can be formed.

REFERENCES

Adams, S. H. (2002). *Communication under stress: Indicators of veracity and deception in written narratives.* Unpublished Ph.D., Virginia Polytechnic Institute and State University, United States -- Virginia.

Adkins, M., Twitchell, D., Burgoon, J. K., & Nunamaker Jr, J. F. (2004). *Advances in automated deception detection in text-based computer-mediated communication.* Paper presented at the Enabling Technologies for Simulation Science VIII, Orlando, FL, USA.

Akehurst, L., Bull, R., Vrij, A., & Kohnken, G. (2004). The effects of training professional groups and lay persons to use criteria-based content analysis to detect deception. *Applied Cognitive Psychology, 18*(7), 877-891.

American heritage dictionary. (1991). (2nd College Edition ed.). Boston: Houghton Mifflin.

Berrien, F. K., & Huntington, G. H. (1943). An exploratory study of pupillary responses during deception. *Journal Of Experimental Psychology, 32*(5), 443-449.

Berry, M. J. A., & Linoff, G. S. (2004). *Data mining techniques* (2 ed.). Indianapolis, Indiana: Wiley Publishing.

Biros, D. P., Hass, M. C., Wiers, K., Twitchell, D., Adkins, M., Burgoon, J. K., et al. (2005). *Task performance under deceptive conditions: Using military scenarios in deception detection research.* Paper presented at the Paper presented at the 38th Annual Hawaii International Conference on System Sciences.

Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Reports, 10*(3), 214-234.

Bond, G. D., & Lee, A. Y. (2005). Language of lies in prison: Linguistic classification of prisoners' truthful and deceptive natural language. *Applied Cognitive Psychology, 19*(3), 313-329.

Bradley, M. T., & Janisse, M. P. (1981). Accuracy demonstrations, threat, and the detection of deception - cardiovascular, electrodermal, and pupillary measures. *Psychophysiology, 18*(3), 307-315.

Buller, D. B., & Burgoon, J. K. (1996). Interpersonal deception theory. *Communication Theory, 6*(3), 203-242.

Buller, D. B., Burgoon, J. K., Buslig, A., & Roiger, J. (1996). Testing interpersonal deception theory: The language of interpersonal deception. *Communication Theory, 6*(3), 268-289.

Burgoon, J. K., Blair, J. P., Qin, T. T., & Nunamaker, J. F. (2003). Detecting deception through linguistic analysis. In *Lecture notes in computer science: Proceedings of intelligence and security informatics: ISI 2003* (Vol. 2665, pp. 91-101).

Burgoon, J. K., Buller, D. B., Guerroro, L. K., Afifi, W., & Feldman, C. (1996). Interpersonal deception: Xii. Information management dimensions underlying deceptive and truthful messages. *Communication Monographs, 63*(1), 50-69.

Burgoon, J. K., Burgoon, M., & Wilkinson, M. (1981). Writing style as a predictor of newspaper readership, satisfaction and image. *Journalism Quarterly, 58*, 225-231.

Burgoon, J. K., Qin, T. T., & Twitchell, D. P. (2006). The dynamic nature of deceptive verbal communication. *Journal of Language & Social Psychology, 25*(1), 1-22.

Cao, J., Crews, J. M., Lin, M., Burgoon, J., & Nunamaker, J. (2003). Designing agent99 trainer: A learner-centered, web-based training system for deception detection. In *Lecture notes in computer science: Proceedings of intelligence and security informatics: ISI 2003.* (Vol. 2665, pp. 358-365).

Cunningham, H. (2002). Gate, a general architecture for text engineering. *Computers and the Humanities, 36*(2), 223-254.

Cunningham, H., Maynard, D., Bontcheva, K., Tablan, V., Ursu, C., Dimitrov, M., et al. (2005, February 2, 2006). Developing language processing components with gate version 3 (a user guide) http://gate.Ac.Uk/sale/tao/index.Html#x1-1710008.4. Retrieved February 15, 2006, from http://gate.ac.uk/sale/tao/index.html#x1-1710008.4

Cutrow, R. J., Parks, A., Lucas, N., & Thomas, K. (1972). The objective use of multiple physiological indices in the detection of deception. *Psychophysiology, 9*(6), 578-588.

Davis, M., Markus, K. A., Walters, S. B., Vorus, N., & Connors, B. (2005). Behavioral cues to deception vs. Topic incriminating potential in criminal confessions. *Law And Human Behavior, 29*(6), 683-704.

DePaulo, B. M. (1992). Nonverbal behavior and self-presentation. *Psychological Bulletin, 111*, 203-243.

DePaulo, B. M., Kirkendol, S. E., Kashy, D. A., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal Of Personality And Social Psychology, 70*(5), 979-995.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129*(1), 74-118.

DePaulo, B. M., Zuckerman, M., & Rosenthal, R. (1980). Humans as lie detectors. *Journal Of Communication, 30*(2), 129-139.

Driscoll, L. N. (1994). A validity assessment of written statements from suspects in criminal investigations using the scan technique. *Police Studies, 17*(4), 77-88.

Ekman, P. (1985). *Telling lies: Clues to deceit in the marketplace, politics, and marriage*. New York: WW Norton & Company.

Ekman, P., & Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry: Journal for the Study of Interpersonal Processes, 32*, 88-105.

Ekman, P., & Friesen, W. V. (1972). Hand movements. *Journal Of Communication, 22*(4), 353-374.

Ekman, P., & Friesen, W. V. (1974). Detecting deception from the body or face. *Journal of Personality and Social Psychology, 29*(3), 288-298.

Ekman, P., & O'Sullivan, M. (1991). Who can catch a liar? *American Psychologist, 46*(9), 913-920.

Ekman, P., O'Sullivan, M., & Frank, M. G. (1999). A few can catch a liar. *Psychological Science, 10*(3), 263-265.

Engelbrecht, A. P., Cloete, I., & Zurada, J. M. (1995). Determining the significance of input parameters using sensitivity analysis. In *From natural to artificial neural computation* (Vol. 930, pp. 382-388). Berlin: Springer-Verlag.

Feeley, T., H., & Young, M., J. . (1998). Humans as lie detectors: Some more second thoughts. *Communication Quarterly, 46*(2), 109-126.

Feeley, T. H., & deTurck, M. A. (1995). Global cue usage in behavioral lie detection. *Communication Quarterly, 43*(4), 420-430.

Feeley, T. H., & deTurck, M. A. (1998). The behavioral correlates of sanctioned and unsanctioned deceptive communication. *Journal of Nonverbal Behavior, 22*(3), 189-204.

Feeley, T. H., & Young, M. J. (2000). Self reported cues about deceptive and truthful communication: The effects of cognitive capacity. *Communication Quarterly, 48*, 101-119.

Fiedler, K., & Walka, I. (1993). Training lie-detectors to use nonverbal cues instead of global heuristics. *Human Communication Research, 20*(2), 199-223.

Ford, E. B. (2006). Lie detection: Historical, neuropsychiatric and legal dimensions. *International Journal Of Law And Psychiatry, 29*(3), 159-177.

Frank, M. G., & Ekman, P. (1997). The ability to detect deceit generalizes across different types of high-stake lies. *Journal of Personality and Social Psychology, 72*(6), 1429-1439.

Frank, M. G., & Ekman, P. (2004). Appearing truthful generalizes across different deception situations. *Journal of Personality and Social Psychology, 86*(3), 486-495.

Frank, M. G., & Feeley, T. H. (2003). To catch a liar: Challenges for research in lie detection training. *Journal Of Applied Communication Research, 31*(1), 58-75.

Friedman, H. S., & Tucker, J. S. (1990). Language and deception. In H. Giles & W. P. Robinson (Eds.), *Handbook of language and social psychology* (pp. 257-270): John Wiley & Sons.

Fuller, C., Biros, D. P., Adkins, M., Burgoon, J., Nunamaker Jr, J. F., & Coulon, S. (2006). Detecting deception in person-of-interest statements. *Lecture Notes in Computer Science, 3975*, 504-509.

Fuller, C., Biros, D. P., Twitchell, D., Burgoon, J., & Adkins, M. (2006, August 4-6, 2006). *An analysis of text-based deception detection tools.* Paper presented at the Twelfth Americas Conference on Information Systems, Acapulco, Mexico.

Gamer, M., Rill, H. G., Vossel, G., & Godert, H. W. (2006). Psychophysiological and vocal measures in the detection of guilty knowledge. *International Journal Of Psychophysiology, 60*(1), 76-87.

George, J. F., & Keane, B. T. (2006). *Deception detection by third party observers*. Paper presented at the Deception Detection Symposium, 39th Annual Hawaii International Conference on System Sciences.,

Godert, H. W., Gamer, M., Rill, H. G., & Vossel, G. (2005). Statement validity assessment: Inter-rater reliability of criteria-based content analysis in the mock-crime paradigm. *Legal And Criminological Psychology, 10*, 225-245.

Granhag, P. r. A., & Stromwall, L. A. (2001). Deception detection: Interrogators' and observers' decoding of consecutive statements. *Journal of Psychology, 135*(6), 603-620.

Grover, S. L. (1993). Lying, deceit, and subterfuge: A model of dishonesty in the workplace. *Organization Science, 4*(3), 478-495.

Grubin, D., & Madsen, L. (2005). Lie detection and the polygraph: A historical review. *Journal Of Forensic Psychiatry & Psychology, 16*(2), 357-369.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis* (5 ed.). Upper Saddle River, New Jersey: Prentice-Hall, Inc.

Hancock, J., Curry, L., Goorha, S., & Woodworth, M. (2004). *Lies in conversation: An examination of deception using automated linguistic analysis.* Paper presented at the Annual Conference of the Cognitive Science Society, Mahwah, NJ.

Hancock, J., Thom-Santelli, J., & Ritchie, T. (2004). *Deception and design: The impact of communication technology on lying behavior.* Paper presented at the SIGCHI conference on Human factors in computing systems, Vienna, Austria.

Hancock, J. T., Curry, L., Goorha, S., & Woodworth, M. (2005). *Automated linguistic analysis of deceptive and truthful synchronous computer-mediated communication.* Paper presented at the 38th Annual Hawaii International Conference on System Sciences.

Hocking, J. E., & Leathers, D. G. (1980). Nonverbal indicators of deception: A new theoretical perspective. *Communication Monographs, 47*(2), 119-131.

Hollien, H., & Harnsberger, J. D. (2006). *Voice stress analyzer instrumentation evaluation* (No. Final Report, CIFA Contract FA 4814-04-0011): IASCP, University of Florida.

Hopkins, C. S., Benincasa, D. S., Ratley, R. J., & Grieco, J. J. (2005). *Evaluation of voice stress analysis technology.* Paper presented at the 38th Annual Hawaii International Conference on System Sciences.

Horvath, F., Jayne, B., & Buckley, J. (1994). Differentiation of truthful and deceptive criminal suspects in behavior analysis interviews. *Journal Of Forensic Sciences, 39*(3), 793-807.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88*(1), 67-85.

Knapp, M. L., Hart, R. P., & Dennis, H. S. (1974). An exploration of deception as a communication construct. *Human Communication Research, 1*, 15-29.

Kohnken, G. (1985). Speech and deception of eyewitnesses: An information processing approach. In F. L. Denmark (Ed.), *Social/ecological psychology and the psychology of women*. North-Holland: Elsevier Science Publishers.

Krauss, R. M., Geller, V., & Olsen, C. (1976). *Modalities and cues in the detection of deception.* Paper presented at the American Psychological Association, Washington, D.C.

Kraut, R. E., & Poe, D. B. (1980). Behavioral roots of person perception: The deception judgments of customs inspectors and laymen. *Journal of Personality and Social Psychology, 39*(5), 784-798.

Lakhani, M., & Taylor, R. (2003). Beliefs about the cues to deception in high- and low-stake situations. *Psychology Crime & Law, 9*(4), 357-368.

Lesce, T. (1990). Scan: Deception detection by scientific content analysis. *Law and Order, 38*(8).

Levine, T. R., Kim, R. K., Park, H. S., & Hughes, M. (2006). Deception detection accuracy is a predictable linear function of message veracity base-rate: A formal test of park and levine's probability model. *Communication Monographs, 73*(3), 243-260.

Mann, S., Vrij, A., & Bull, R. (2002). Suspects, lies, and videotape: An analysis of authentic high-stake liars. *Law And Human Behavior, 26*(3), 365-376.

Mann, S., Vrij, A., & Bull, R. (2004). Detecting true lies: Police officers' ability to detect suspects' lies. *Journal Of Applied Psychology, 89*(1), 137-149.

Masip, J., Sporer, S. L., Garrido, E., & Herrero, C. (2005). The detection of deception with the reality monitoring approach: A review of the empirical evidence. *Psychology, Crime & Law, 11*(1), 99-122.

McCornack, S. A. (1992). Information manipulation theory. *Communication Monographs, 59*(1), 1-16.

Meissner, C. A., & Kassin, S. M. (2002). "He's guilty!": Investigator bias in judgments of truth and deception. *Law & Human Behavior, 26*(5), 469-480.

Miller, G. R., & Stiff, J. B. (1993). *Deceptive communication*. Newbury Park: Sage Publications.

National Research Council, (Ed.). (2003). *The polygraph and lie detection*. Washington, DC: The National Academies Press.

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and social psychology bulletin, 29*(5), 665-675.

Park, H. S., & Levine, T. R. (2001). A probability model of accuracy in deception detection experiments. *Communication Monographs, 68*(2), 201-210.

Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Pennebaker, J. W., & Francis, M. E. (2001). Linguistic inquiry and word count: Liwc 2001. Mahwah, NJ: Erlbaum Publishers.

Pollina, D. A., Dollins, A. B., Senter, S. M., Krapohl, D. J., & Ryan, A. H. (2004). Comparison of polygraph data obtained from individuals involved in mock crimes and actual criminal investigations. *Journal Of Applied Psychology, 89*(6), 1099-1105.

Porter, S., & Yuille, J. C. (1995). Credibility assessment of criminal suspects through statement analysis. *Psychology Crime & Law, 1*(4), 319-331.

Porter, S., & Yuille, J. C. (1996). The language of deceit: An investigation of the verbal clues to deception in the interrogation context. *Law And Human Behavior, 20*(4), 443-458.

Qin, T. T., Burgoon, J., Blair, J. P., & Nunamaker Jr, J. F. (2005). *Modality effects in deception detection and applications in automatic-deception-detection.* Paper presented at the Hawaii International Conference on System Sciences.

Rice, B. (1978). The new truth machines. *Psychology Today, 12*(1), 61-64, 67,72, 74, 77-78.

Sarle, W. (2004). What are cross-validation and bootstrapping? , 2005, from http://www.faqs.org/faqs/aifaq/neural-nets/part3/section-12.html.

Sporer, S. L. (1997). The less travelled road to truth: Verbal cues in deception detection in accounts of fabricated and self-experienced events. *Applied Cognitive Psychology, 11*, 373-397.

Sporer, S. L. (2004). Reality monitoring and detection of deception. In *The detection of deception in forensic contexts* (pp. 64-102). New York: Cambridge University Press.

Sporer, S. L., & Schwandt, B. (2006). Paraverbal indicators of deception: A meta-analytic synthesis. *Applied Cognitive Psychology, 20*, 421-446.

Stromwall, L. A., Hartwig, M., & Granhag, P. A. (2006). To act truthfully: Nonverbal behaviour and strategies during a police interrogation. *Psychology Crime & Law, 12*(2), 207-219.

Twitchell, D., Biros, D. P., Forsgren, N., Burgoon, J., & Nunamaker Jr, J. F. (2005). *Assessing the veracity of criminal and detainee statements: A study of real-world data.* Paper presented at the 2005 International Conference on Intelligence Analysis.

Twitchell, D., Jensen, M. L., Burgoon, J. K., & Nunamaker, J. F., Jr. (2004). *Detecting deception in secondary screening interviews using linguistic analysis.* Paper presented at the The 7th International IEEE Conference on Intelligent Transportation Systems, 2004

Undeutsch, U., & Yuille, J. C. (1989). The development of statement reality analysis In *Credibility assessment.* (pp. 101-119): Kluwer Academic/Plenum Publishers.

Vrij, A. (1993). Credibility judgments of detectives: The impact of nonverbal behavior, social skills, and physical characteristics on impression formation. *Journal of Social Psychology, 133*(5), 601-610.

Vrij, A. (1995). Behavioral-correlates of deception in a simulated police interview. *Journal Of Psychology, 129*(1), 15-28.

Vrij, A. (2000). *Detecting lies and deceit: The psychology of lying and the implications for professional practice.* New York: John Wiley & Sons.

Vrij, A. (2005). Criteria-based content analysis - a qualitative review of the first 37 studies. *Psychology Public Policy And Law, 11*(1), 3-41.

Vrij, A., Edward, K., & Bull, R. (2001). People's insight into their own behaviour and speech content while lying. *British Journal Of Psychology, 92*, 373-389.

Vrij, A., Edward, K., Roberts, K. P., & Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal behavior, 24*(4), 239-263.

Vrij, A., & Mann, S. (2001a). Telling and detecting lies in a high-stake situation: The case of a convicted murderer. *Applied Cognitive Psychology, 15*(2), 187-203.

Vrij, A., & Mann, S. (2001b). Who killed my relative? Police officers' ability to detect real-life high-stake lies. *Psychology Crime & Law, 7*(2), 119-132.

Vrij, A., & Mann, S. (2004). Detecting deception: The benefit of looking at a combination of behavioral, auditory, and speech content related cues in a systematic manner. *Group Decision and Negotiation, 13*(1), 61-79.

Vrij, A., Mann, S., & Fisher, R. (2006). An empirical test of the behaviour analysis interview. *Law & Human Behavior, 30*(3), 329-345.

Vrij, A., Semin, G. R., & Bull, R. (1996). Insight into behavior displayed during deception. *Human Communication Research, 22*(4), 544-562.

Watson, K. W., & Ragsdale, J. D. (1981). Linguistic indices of truthful and deceptive responses to employment interview questions. *Journal Of Applied Communication Research, 9*(2), 59-71.

Weiss, S. M., & Kulikowski, C. A. (1991). *Computer systems that learn: Classification and prediction methods from statistics, neural nets, machine learning, and expert systems*. San Mateo, California: Morgan Kaufman Publishers, Inc.

Whissell, C. (1989). Whissell's dictionary of affect in language: Technical manual and user's guide.

Wiener, M., & Mehrabian, A. (Eds.). (1968). *Language within language: Immediacy, a channel in verbal communication*. New York: Appleton-Century-Crofts.

Witten, I. H., & Frank, E. (2000). *Data mining: Practical machine learning tools and techniques with java*. San Francisco: Morgan Kaufman.

Zhou, L., Burgoon, J. K., Nunamaker, J., Jay F., & Twitchell, D. P. (2004). Automated linguistics based cues for detecting deception in text-based asynchronous computer-mediated communication: An empirical investigation. *Group Decision and Negotiation, 13*(1), 81-106.

Zhou, L., Burgoon, J. K., & Twitchell, D. P. (2003, June 2-3, 2003). *A longitudinal analysis of language behavior of deception in e-mail.* Paper presented at the First NSF/NIJ Symposium on Intelligence and Security Informatics (ISI 2003).

Zhou, L., Burgoon, J. K., Twitchell, D. P., Qin, T. T., & Nunamaker, J. F. (2004). A comparison of classification methods for predicting deception in computer-mediated communication. *Journal Of Management Information Systems, 20*(4), 139-165.

Zhou, L., Twitchell, D. P., Qin, T. T., Burgoon, J. K., & Nunamaker, J. F., Jr. (2003). *An exploratory study into deception detection in text-based computer-mediated communication.* Paper presented at the 36th Annual Hawaii International Conference on System Sciences, Big Island, Hawaii.

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 14, pp. 1-59). New York: Academic Press.

Zuckerman, M., & Driver, R. E. (1987). Telling lies: Verbal and nonverbal correlates of deception. In *Multichannel intergration of nonverbal behavior*. Hillsdale, NJ: Erlbaum.

APPENDICES

Appendix A

Severity Rankings

Please rate the following incident types on a scale from 1 to 5, as shown below, where 1 is the least severe and 5 is the most severe, in terms of punishment:

1	2	3	4	5
Least Severe				Most Severe

Type Of Incident	Least Severe				Most Severe
Domestic Disturbance/Dispute	1	2	3	4	5
Witness to Domestic Disturbance	1	2	3	4	5
Animal Control	1	2	3	4	5
Sexual Harassment	1	2	3	4	5
Harassment	1	2	3	4	5
Single Vehicle Fender Bender	1	2	3	4	5
Decal Sticker Lost/Stolen	1	2	3	4	5
Witness to Parking Ticket	1	2	3	4	5
Credit Card Theft	1	2	3	4	5
Report of Expired Vehicle Registration	1	2	3	4	5
Financial Irresponsibility	1	2	3	4	5
Report of Gas Theft	1	2	3	4	5
Report of Belligerent Suspect	1	2	3	4	5
Misuse of Government Credit Card	1	2	3	4	5
Theft	1	2	3	4	5

Type Of Incident	Least Severe				Most Severe
Assault	1	2	3	4	5
Witness to DWI	1	2	3	4	5
Drunk on Station	1	2	3	4	5
Threat	1	2	3	4	5
Shoplifting	1	2	3	4	5
Loss of Government Equipment	1	2	3	4	5
Destruction of Government Property/Disobeying Lawful order	1	2	3	4	5
DWI	1	2	3	4	5
Drugs in Vehicle	1	2	3	4	5
Suspected Drug Use	1	2	3	4	5
False Witness Statement	1	2	3	4	5
Gun on Base	1	2	3	4	5
BB Gun Incident	1	2	3	4	5
False Accusation of Assault	1	2	3	4	5
Road Rage	1	2	3	4	5
Minor in Possession	1	2	3	4	5
Purchasing Alcohol for a Minor	1	2	3	4	5
Open Container	1	2	3	4	5
Insubordination	1	2	3	4	5
Credit Card Fraud	1	2	3	4	5
Unauthorized Vehicle Use	1	2	3	4	5
Inappropriate Material on Government computer	1	2	3	4	5
Vandalism	1	2	3	4	5
Forgery	1	2	3	4	5
Witness to Vandalism	1	2	3	4	5
Sexual act with a minor	1	2	3	4	5

Type Of Incident	Least Severe				Most Severe
Drug Abuse	1	2	3	4	5
Failure to Obey Direct Order	1	2	3	4	5
Leaving Station without permission	1	2	3	4	5
Unauthorized use of Government Vehicle/Fleeing scene	1	2	3	4	5
Vehicle Break-in	1	2	3	4	5
Driving with a suspended license	1	2	3	4	5
Witness to Assault	1	2	3	4	5
Safety Violation	1	2	3	4	5
Witness to Underage Drinking	1	2	3	4	5
Witness to arson	1	2	3	4	5
Destruction of Government Records	1	2	3	4	5

Appendix B

Statement Transcription Procedures

I. Prepare Written Statements

   a. Black out personal information (Name, SSN, etc.). Next to the information that has been blacked out, indicate what type of information has been blacked out.

   b. Systematically replace names identified in the statement with a dummy name. For example, each instance of the fourth male mentioned is replace with the name “John”.

   c. Label the statement as Truthful, Deceptive or Unknown, and label with the gender of the person of interest.

II. Transcribe Written Statements

   a. Open Notepad or WordPad

   b. Type statement exactly how it is written

      i. Match Case

      ii. Match punctuation

      iii. Match spelling

      iv. Replace the names from the written statement with the dummy names as described above.

      v. For corrections made by person of interest: ignore the initials used to verify the corrections were made by the person of interest and put whatever was marked out in brackets, “[ ].” This will allow for an automated extraction or classification technique to be created and used by the information system.

vi. For any words that are illegible (either partially or entirely), place these words in curly braces. For any letters that can't be read, use percent (%) as a placeholder. For example, if the letter x in the word text can't be made out, this should be transcribed as curly brace te%t curly brace.

vii. For any information that is marked out by someone other than the person of interest (for example, by law enforcement personnel):

a. If type of information is known, replace with similar information. For example if a social security number was blacked out, replace with 123-45-6789.

b. If the type of information blacked out is unknown, indicate this by typing |x| to indicate that there was information on the statement of an unknown type blacked out.

viii. Two-person statements (those that include Q&A segments)should not be included in the data set.

III. Saving Typed Statements

a. Save statements with the last name of who made the transcription, gender of the person of interest (0=unknown, 1=male, 2=female), True or False, a 5-digit number, a letter or letters to indicate where the statement was collected, and ".txt". (i.e. the first false statement from a male collected at Generic Base will be: “Smith1False00001G.txt” and the third truthful statement by a female from Generic Base is: “Smith2True00003G.txt”

Appendix C

IRB Documentation

Oklahoma State University Institutional Review Board

Request for Determination of Non-Human Subject or Non-Research

D. Are data/specimens <u>received</u> by the Investigator with identifiable private information?
☒ No ☐ Yes

E. Are the data/specimen(s) coded such that a link exists that could allow the data/specimen(s) to be re-identified?
☒ No ☐ Yes
If "*Yes*," is there a written agreement that prohibits the PI and his/her staff access to the link?
☐ No ☐ Yes

6. Signatures

Signature of PI ____________ **Date** 9/18/06

Signature of Faculty Advisor David P. Biros **Date** 18 Sept 06
(If PI is a student)

☒ Based on the information provided, the OSU-Stillwater IRB has determined that this research **does not** qualify a human subject research as defined in 45 CFR 46.102(d) and (f) and **is not subject to oversight by the OSU IRB.**

☐ Based on the information provided, the OSU-Stillwater IRB has determined that this research **does** qualify as human subject research and **submission of an application for review by the IRB is required.**

Sue C Jacobs
Dr. Sue C. Jacobs, IRB Chair

9/20/06
Date

Revision Date: 04/2006

5 of 5

VITA

Christie Marlene Fuller

Candidate for the Degree of Doctor of Philosophy

Dissertation: HIGH-STAKES, REAL-WORLD DECEPTION: AN EXAMINATION OF THE PROCESS OF DECEPTION AND DECEPTION DETECTION USING LINGUISTIC-BASED CUES

Major Field: Business Administration

Biographical:

Education: Graduated from Hays High School, Hays, Kansas in May 1994; received Bachelor of Science degree in Management from Kansas State University, Manhattan, Kansas in December 1998; received Master of Business Administration degree from Fort Hays State University, Hays, Kansas in December 2001. Completed the requirements for the Doctor of Philosophy degree with a major in Business Administration at Oklahoma State University in May, 2008.

Experience: Employed as a graduate assistant at Fort Hays State University, College of Business and Leadership, 2000 to 2001. Employed as an Instructor at Fort Hays State University, Computer Information Systems Department, 2002 to 2003. Employed as a graduate associate by Oklahoma State University, Department of Management Science and Information Systems, 2003 to 2007. Employed by Oklahoma State University-Tulsa as a visiting assistant professor, 2008.

Professional Memberships: Association for Information Systems, Decision Sciences Institute, Phi Kappa Phi Honor Society