

Who Should You Trust? Discriminating Genuine from Deceptive Eyewitness Accounts

Charles A. Morgan III^{1,*}, Yaron Rabinowitz², Beau Palin³ and Kirk Kennedy⁴

¹National Security Program, Henry C. Lee College of Criminal Justice & Forensic Sciences, University of New Haven, West Haven, CT, United States

²U.S. Navy, Camp Lejeune, NC, United States

³Jonathan Edwards College, Yale University, New Haven, CT, United States

⁴FBI Critical Incident Response Group, Behavioral Research and Instruction Unit, Quantico, VA, United States

Abstract: In this study, we tested whether modified cognitive interviewing (MCI) is a valid method for distinguishing between genuine and deceptive human eyewitness accounts.

102 healthy military personnel were the participants of this study. 54 were assigned to a manual task condition and 48 to a cognitive task condition. Of the 54 assigned to the manual task, 17 truly performed the task and were truthful when interviewed about their activities; 18 performed the task and, when interviewed denied having performed the task; 19 read the instructions regarding the manual task and when interviewed falsely claimed to have performed the task.

Of the 48 participants assigned to the cognitive task, 20 performed the task and reported truthfully about their activities; 13 performed the task and denied having participated in the task; 15 read the instructions about the cognitive task and when interviewed claimed to have actually performed the task. The transcripts of interviews were rated by individuals trained in cognitive interviewing; forensic speech variables (response length, unique word count and type-token ratio (TTR)) were coded from transcripts.

Human rater judgments and computer-based speech analysis performed better than chance; computer based judgments were superior to the human judgments (i.e., 80% vs. 62%, respectively). Speech content variables derived from MCI differed significantly, and in different ways, between the truthful and false claimant participants and also between the truthful and denial type participants. MCI derived statement analysis methods are a scientifically valid method that can be used by professionals tasked with distinguishing between true claims, false claims and denials.

Keywords: Credibility assessment, deception detection, eyewitness memory, forensic psychiatry, security.

INTRODUCTION

At present, U.S. government officials have few objective, scientifically valid, ways to determine whether people who offer (or deny knowing) information of relevance to national security are being genuine or deceptive. Indeed, the majority of scientific studies demonstrate that professional and laypersons alike perform below levels of chance when attempting to detect deception (Morgan *et al.*, 2013; Vrij & Akehurst, 1998; Bond *et al.*, 1985; Government-Report, 2009). This has been demonstrated to be true among many professional groups including police officers, judges,

psychiatrists, university students and agents from government law enforcement agencies (Ekman & O'Sullivan, 1991; DePaulo & Pfiefer, 1986; Newman *et al.*, 2005; Vrij & Akehurst, 1998 for full review). Thus, the further development of methods for assessing human deception could potentially help government officials to improve the validity and efficiency of intelligence assessments.

Over the past ten years, a significant body of literature has emerged that demonstrates that speech content analysis – when derived through the use of a specific interviewing technique known as modified cognitive interviewing (MCI) – can be used to detect deception at rates significantly above other current approaches used by professionals. Specifically, analysis of speech content – not voice stress – has produced detecting deception rates at or above 82% – a rate that is significantly higher than that achieved by polygraphy (i.e. 65%) or by chance (i.e. 50%) (Vrij, 2008 for full review, Colwell *et al.*, 2002; 2007; Morgan *et al.*, 2007; 2008; 2008b; 2009; 2013; 2014).

*Address correspondence to this author at the National Security, Henry C. Lee College of Criminal Justice, University of New Haven, 234 Church Street #301, New Haven, CT 06510; Tel: 203 932 1154; Emails : CMorgan@newhaven.edu; camorgan3rd@gmail.com

§Study was conducted when CAM was with the National Center for PTSD, VA CT Yale University School of Medicine, New Haven CT, United States

While the data from our previous studies are promising, the majority of studies on deception have focused on detecting lies of omission. Few studies have focused on detecting lies of fabrication (i.e., false claims) (Morgan *et al.*, 2004), (Morgan, *et al.* 2007; Harnesberger *et al.* 2009; Hollien *et al.*, 2008; Morgan *et al.*, 2007; 2008; 2008b; 2009). Further, at present no studies have examined whether this methodology could effectively be used to discriminate truthful claims from denials or false claims in persons questioned about the same task or activity. Understanding whether alterations in speech content vary significantly between truthful claims, false claims or denials would clarify how well the method can be applied by professionals in the real world. The present study was designed to examine: A) whether MCI derived speech content could be used to distinguish truthful accounts from false claims or denials in people who performed a Manual or a Cognitive activity; and B) whether the patterns observed for the two different activities would differ.

METHODS

Subjects: The participants of this study were 102 active duty Marines stationed at Camp Lejeune, NC. Each was given an oral briefing about the project and subsequently provided written informed consent prior to participation in the study. All participants indicated that they understood that participation was optional and their decision about whether to participate would not be relayed back to, or have any bearing on their status in the command. All participants understood this project was sponsored by the HIG/FBI and not the US Navy or Marine Corps.

DESIGN

This study design consisted of two phases (see flow chart below): In Phase One participants engaged in, or only read about, a manual or a cognitive task; in Phase Two participants were interviewed about their claimed activity (i.e., their activity during Phase One) by interviewers who were blind to the status of the participant (i.e. they did not know whether or not the participant had truly engaged in the manual or cognitive tasks during Phase One). When interviewing participants during Phase Two, the interviewers used a Modified Cognitive Interview (MCI) technique (Morgan *et al.*, 2007; 2013; 2014). Transcripts from the interviews with participants were rated by human raters and also used to generate speech content variables for computer analysis.

Phase One: Task Exposure

Depending on the randomization system, participants in Phase One engaged in, or only read about one of two different tasks: a manual task (see below for a more detailed description) or a cognitive task (see below for more detailed description). All truthful persons completed either the manual task or the cognitive task. All deceptive participants assigned to the “denial” group also completed either the manual or the cognitive task. Deceptive participants assigned

to the “fabrication” group were only permitted to read the instructions for one (but not both) of the tasks.

The manual task involved having participants learn how to operate a remote controlled toy helicopter. They would then learn to fly it through an obstacle course set up in an indoor room. Although learning to fly the helicopter may seem to be a relatively simple task, it is complex and requires practice. Participants were given a one hour period of time in which to practice in order to achieve a successful completion of the obstacle course.

The cognitive task involved participants in a series of timed trials during which they had to make use of a set of shapes (i.e. using a commercially available game called Tangoes) to construct an image that matched the figure shown to them by the instructor administering the task. This game is mentally challenging, and as observed in previous military exercises, requires significant mental effort for the participant to complete in an accurate manner. The task was considered “complete” when participants completed the task or when the time expired (which ever came first).

After completing their task, participants assigned to the Truthful condition were told that they would be interviewed about how they had spent their time. They were instructed to respond openly and honestly about the nature of their activities. They were given 10 minutes to collect their thoughts prior to the interview. Conversely, after completing their task, participants assigned to the “denial” deceptive condition were told that they could not report on their activities and were instructed to lie by reporting that they had spent their time reading a magazine and discussing it with a colleague. Each was handed a copy of a popular magazine that was taped shut. Participants were told to examine the cover but not the contents of the magazine. After being given approximately 10 minutes to collect their thoughts, the “denial” participants participated in the cognitive interview.

Participants assigned to the “false claim” deceptive condition were given written detailed instructions about either the manual or the cognitive task. Each was given 30 minutes to study the materials. They were told when given the materials that they would have to lie and claim that they had actually performed the task when interviewed.

NOTE: The act of tasking participants in the “denial” condition to lie about their time and to also claim that they had read a magazine created a situation in which “denial” participants not only denied knowledge of the actual task, but also, told a lie of fabrication (i.e., they claimed to have read a magazine and to have discussed it with a colleague). This was a deliberate design choice based upon previous findings that simple lies of omission (i.e., a situation in which a person simply denies any knowledge about the events in question) often resulted in impoverished reporting which, in turn, led to an easier detection of deception. By asking individuals to fabricate a story, we hoped to increase participants’ verbal productivity, and, in doing so make it more difficult to detect deception by word count alone. These participants are still referred to as “deniers” in keeping with the face validity terms used in the field of detecting

deception: they claim not to know anything about the events in question.

Phase Two: The Modified Cognitive Interview (MCI)

Each participant was interviewed approximately 10 minutes after completing their manual, cognitive or reading tasks. As in our previous studies, the interviews with participants occurred in a well lighted room and were recorded. These recordings were ultimately transcribed. Each transcript was edited so as to create two new files: one file contained the speech/statements made by the interviewee; the second file contained only speech/statements made by the interviewer; Interviewee (participant) transcripts were used to calculate the key Speech Content variables that would be used in the actuarial database: Response Length [i.e. total words uttered by participant], Unique Word Count [total number of unique words spoken] and the Type-Token Ratio (TTR) [i.e. the ratio of Unique words/Response Length]. Based on previous data showing that some phases of the MCI may be more useful than others at discriminating between genuine and deceptive responses, we created these variables for each phase of the MCI. Finally, in order to assess whether the best discrimination would occur when using the interview as a whole, we created these three variables from the entire speech/statement set acquires during the cognitive interview for a specific task (manual or cognitive).

SUMMARY FLOW CHART OF STUDY DESIGN

PHASE ONE

Participants Recruited for the Study

Participants Randomized to ONE of Four Possible Conditions:

- 1) Genuine Participation in a Manual Task.
- 2) Genuine Participation in a Cognitive Task.
- 3) ONLY reading about the Manual Task.
- 4) ONLY reading about the Cognitive Task.

PHASE TWO

Participants from Groups 1 & 2 were randomized to TRUTHFUL or DENIAL conditions.

Participants from Groups 3 & 4 were assigned to the Deceptive FABRICATOR condition.

NOTE: This design permitted the assessment of genuine claims; the assessment of denials from individuals who had actually performed the tasks; and an assessment of deceptive individuals who claimed to have actually performed the tasks (false claimants).

STATISTICAL ANALYSIS

Transcript Rater Judgments

Three college students who received training in cognitive interviewing methods by the PI of this study (CAM) served

as the raters. Each independently reviewed the typed interview transcripts of the MCI. After reading a transcript, the raters rendered a judgment about a participants' status (Truthful/Deceptive) based on whether raters assessed the mnemonic prompt phase of the MCI to have elicited significantly more detail than the initial "detail" prompt. Rater judgments were contrast coded with deceptive ratings coded as a "1" and genuine ratings coded as a "0". Individual Cross-tab analyses were performed using the variables Genuine Status (i.e. the true assignment of the participant) and each individual's Judgment Scores.

Forensic Statement Analysis

In order to assess whether Response Length, Unique Word Count, TTR for the Manual task differed between truthful and deceptive participants we performed General Linear Model Multivariate Analyses of Variance using Group (Truthful, False Claim, Denial) as the independent variable and Speech Content (i.e. TTR, Response Length and Unique Word count from each prompt of the MCI) as the dependent variables. (SPSS 2010) Tukey post hoc tests were used to evaluate how speech content variables differed amongst the three groups (Truthful, False Claim, Denial). Similar analyses were performed for the data set related to the Cognitive task.

Receiver Operating Characteristic (ROC) curves were generated to evaluate which variables best discriminated between the three groups of participants. (Metz 1978; Ahtham, 1973). We then performed binary logistic regressions with the most useful speech content variables to calculate the classification accuracy of the models. (SPSS 2010)

RESULTS

Rater Accuracy: The human raters performed better than chance at distinguishing true from false accounts. Cross tab analyses indicated that the classification accuracies for each of the three raters differed significantly from chance. Rater #1's accuracy was 62% (78% correct for truthful accounts; 45% correct for deceptive accounts); Rater #2's accuracy was 61% (74% correct for truthful accounts; 48% correct for deceptive accounts); Rater #3's accuracy was 62% (76% correct for truthful accounts; 47% correct for deceptive accounts).

Forensic Statement Analysis Results

Analyses of the MCI derived speech content variables [Type-Token Ratio (TTR), Response Length (RL) and Unique Words (UW)] for the Manual and Cognitive Tasks are noted below in Tables 1 and 2, respectively.

Manual Task

General Linear Model Multivariate Analyses of Variance for the Manual Task indicated the presence of significant differences between the 3 groups of participants on the speech content variables. With respect to the variable TTR, significant differences were noted for: Prompt One (F (2,1) =

Table 1. Speech variables for truthful, false claims and denial participants (Manual Task).

| | Truthful | False Claims | Deceptive |
|--|---------------------|--------------------------|-----------------------------|
| Prompt One [Detailed Account] | | | |
| TTR | 0.42 (SD = 0.1)* | 0.39 (SD = 0.1)* | 0.57 (SD = .14) γ + |
| RL | 405.0 (SD = 212)* | 406.0 (SD = 246)* | 184.6 (SD = 107) γ + |
| UW | 153.9 (SD = 60)* | 138.2 (SD = 57)* | 91.9 (SD = 42) γ + |
| Prompt Two [Visual Prompt] | | | |
| TTR | 0.49 (SD = 0.1)* | 0.69 (SD = 0.2) γ | .62 (SD = .01) γ |
| RL | 239.5 (SD = 132)*+ | 95.8 (SD = 77) γ | 99.6 (SD = 46) γ |
| UW | 104.8 (SD = 40)* | 54.4 (SD = 30) γ | 57.2 (SD = 18) γ |
| Prompt Three [Auditory Prompt] | | | |
| TTR | 0.50 (SD = .15)* | 0.62 (SD = 0.2) | .68 (SD = .14) γ |
| RL | 219.7 (SD = 134)*+ | 98.1 (SD = 65) | 82.7 (SD = 78) γ |
| UW | 94.6 (SD = 42)* | 55.3 (SD = 31) | 46.9 (SD = 31) γ |
| Prompt Four [Emotional Prompt] | | | |
| TTR | 0.61 (SD = .14) | 0.80 (SD = 0.2)* | .69 (SD = .11) + |
| RL | 136.9 (SD = 43)*+ | 37.1 (SD = 29)* | 75.1 (SD = 48) γ + |
| UW | 70.5 (SD = 46)+ | 25.9 (SD = 18)* | 47.7 (SD = 25)+ |
| Prompt Five [Temporal Prompt] | | | |
| TTR | 0.46 (SD = .11)* | 0.57 (SD = .11) | .54 (SD = .12) γ |
| RL | 250.5 (SD = 139)*+ | 110.3 (SD = 63) γ | 127.9 (SD = 81) γ |
| UW | 102.0 (SD = 36)* | 56.6 (SD = 25) | 63.7 (SD = 32) γ |
| Prompt Six [Mistakes Prompt] | | | |
| TTR | 0.80 (SD = .17) | 0.82 (SD = .16) | .89 (SD = .13) |
| RL | 36.9 (SD = 43) | 24.7 (SD = 26) | 20.3 (SD = 25) |
| UW | 24.8 (SD = 22) | 17.3 (SD = 17) | 15.4 (SD = 18) |
| Total Interview | | | |
| TTR | 0.27 (SD = .07)* | 0.32 (SD = .07)* | .37 (SD = .06) γ + |
| RL | 1288.5 (SD = 132)*+ | 771.9 (SD = 354) | 590.1 (SD = 256) γ + |
| UW | 311.8 (SD = 105)* | 224.1 (SD = 78) | 202.5 (SD = 67) γ + |
| Prompt Total [Sum of Prompts 2-6] | | | |
| RL | 846.5 (SD = 461)*+ | 341.2 (SD = 174) | 385.4 (SD = 180) |
| Detailed Account [Prompt 1] to Prompt Ratio [2-6] | | | |
| RL | 2.2 (SD = 0.8) | 1.1 (SD = 0.9)* | 3.0 (SD = 2.4)+ |

*Variable differs significantly from Denial Group.

+Variable differs significantly from False Claims Group.

 γ Variable differs significantly from Truthful Group.

11.7; $p < 0.001$), Prompt Two ($F(2,1) = 8.6$; $p < 0.001$); Prompt Three ($F(2,1) = 5.2$; $p < 0.009$); Prompt Four ($F(2,1) = 8.3$; $p < 0.001$); Prompt Five ($F(2,1) = 4.4$; $p < 0.04$), and for the Full Interview ($F(2,1) = 9.3$; $p < 0.001$).

Significant differences between the Groups of Participants were also noted for the speech content variable RL in response to Prompt One ($F(2,1) = 7.2$; $p < 0.002$); Prompt Two ($F(2,1) = 14$; $p < 0.000$); Prompt Three ($F(2,1) = 10.3$; $p < 0.001$); Prompt Four ($F(2,1) = 7.8$; $p < 0.001$).

Table 2. Speech variables for truthful, false claims and denial participants (Cognitive Task).

| | <u>Truthful</u> | <u>False Claims</u> | <u>Deceptive</u> |
|--|--------------------|---------------------------|---------------------------|
| Prompt One [Detailed Account] | | | |
| TTR | 0.46 (SD = 0.1)* | 0.49 (SD = 0.1) | 0.56 (SD = .09) γ |
| RL | 351.7 (SD = 245)* | 249.4 (SD = 177) | 158.6 (SD = 93) γ |
| UW | 137.3 (SD = 67)* | 107.1 (SD = 58) | 81.3 (SD = 38) γ |
| Prompt Two [Visual Prompt] | | | |
| TTR | 0.49 (SD = 0.1)+* | 0.66 (SD = 0.2) γ | .67 (SD = .12) γ |
| RL | 257.5 (SD = 187)+* | 98.3 (SD = 82) γ | 75.2 (SD = 41) γ |
| UW | 109.4 (SD = 54)+* | 54.2 (SD = 34) γ | 57.2 (SD = 18) γ |
| Prompt Three [Auditory Prompt] | | | |
| TTR | 0.53 (SD = .16)+* | 0.69 (SD = .12) γ | .65 (SD = .13) γ |
| RL | 210.6 (SD = 132)+* | 80.5 (SD = 63) γ | 88.4 (SD = 78) γ |
| UW | 93.6 (SD = 44)+* | 49.7 (SD = 25) γ | 51.2 (SD = 32) γ |
| Prompt Four [Emotional Prompt] | | | |
| TTR | 0.66 (SD = .17) | 0.73 (SD = .15) | .69 (SD = 14) |
| RL | 124.1 (SD = 100)+ | 64.7 (SD = 54) γ | 68.9 (SD = 49) |
| UW | 67.0 (SD = 43)+ | 41.4 (SD = 26) γ | 43.0 (SD = 25) |
| Prompt Five [Temporal Prompt] | | | |
| TTR | 0.51 (SD = .12)* | 0.58 (SD = .09) | .60 (SD = .14) γ |
| RL | 176.6 (SD = 99)+ | 100.3 (SD = 44) γ | 117.0 (SD = 66) |
| UW | 80.0 (SD = 28)+ | 55.3 (SD = 18) γ | 63 (SD = 30) |
| Prompt Six [Mistakes Prompt] | | | |
| TTR | 0.79 (SD = .18)+* | 0.53 (SD = .46) γ | .93 (SD = .12) γ |
| RL | 38.6 (SD = 45)+ | 11.7 (SD = 16) γ | 14.1 (SD = 18) |
| UW | 24.2 (SD = 23)+ | 9.0 (SD = 13) γ | 11.0 (SD = 13) |
| Total Interview | | | |
| TTR | 0.28 (SD = .07)+* | 0.36 (SD = .07) γ | .37 (SD = .06) γ |
| RL | | | |
| UW | 284.5 (SD = 110)+* | 197.0 (SD = 75) γ | 180.5 (SD = 69) γ |
| Prompt Total [Sum of Prompts 2-6] | | | |
| RL | 807.4 (SD = 478)+* | 355.5 (SD = 215) γ | 363.6 (SD = 198) γ |
| Detailed Account [Prompt 1] to Prompt Ratio [2-6] | | | |
| RL | .44 (SD = .20)+ | .80 (SD = .73) γ | .45 (SD = .17) |

*Variable differs significantly from Denial Group.

+Variable differs significantly from False Claims Group.

 γ Variable differs significantly from Truthful Group.

Prompt Five ($F(2,1) = 10.1$; $p < 0.001$) and the Full Interview ($F(2,1) = 11.5$; $p < 0.000$). Finally, significant differences between the Groups of Participants emerged for the speech content variable UW for: Prompt One ($F(2,1) = 6.2$;

$p < 0.004$); Prompt Two ($F(2,1) = 14.3$; $p < 0.000$); Prompt Three ($F(2,1) = 9.1$; $p < 0.001$); Prompt Four ($F(2,1) = 8.4$; $p < 0.001$); Prompt Five ($F(2,1) = 10.5$; $p < 0.000$) and the Full Interview ($F(2,1) = 8.3$; $p < 0.001$).

Table 3. Area under the curve associated with manual task speech variables.

| | <u>Area</u> | <u>Asymptotic Sig.</u> |
|------------------|-------------|------------------------|
| Prompt One UW | .70 | p< 0.05 |
| Total RL | .80 | p< 0.001 |
| Prompt Total | .85 | p<0.001 |
| Prompt Ratio .70 | p<0.04 | |

Table 4. Area under the curve associated with cognitive task speech variables.

| | <u>Area</u> | <u>Asymptotic Sig.</u> |
|------------------|-------------|------------------------|
| Prompt One RL | .70 | p< 0.02 |
| Total RL | .80 | p< 0.001 |
| Prompt Total | .82 | p<0.001 |
| Prompt Ratio .40 | p<0.09 | |

Cognitive Task

General Linear Model Multivariate Analyses of Variance for the Cognitive Task indicated the model was significant and that the 3 groups of participants differed significantly in many of the speech content variables. With respect to TTR, significant differences were noted for: Prompt Two (F (2,1) = 7.7; p<0.001), Prompt Three (F (2,1) = 6.7; p<0.003); Prompt Six (F (2,1) = 7.0; p<0.002), and for the Full Interview (F (2,1) = 8.7; p<0.001).

Significant differences between the Groups of Participants were also noted for the speech content variable RL in response to Prompt One (F (2,1) = 4.0; p<0.02); Prompt Two (F (2,1) = 9.8; p<0.000); Prompt Three (F (2,1)

= 9.2; p<0.000); Prompt Four (F (2,1) = 3.3; p<0.04) Prompt Five (F (2,1) = 4.7; p<0.01); Prompt Six (F (2,1) = 3.8; p<0.03) and the Full Interview (F (2,1) = 8.4; p<0.001).

Finally, significant differences between the Groups of Participants were noted for the speech content variable UW for: Prompt One (F (2,1) = 3.7; p<0.03); Prompt Two (F (2,1) = 12; p<0.000); Prompt Three (F (2,1) = 8.7; p<0.001); Prompt Four (F (2,1) = 3.1; p<0.05); Prompt Five (F (2,1) = 4.0; p<0.02); Prompt Six (F (2,1) = 3.7; p<0.03) and the Full Interview (F (2,1) = 6.6; p<0.003).

The primary ways in which the Truthful, False Claims and Denial groups for the Manual Task and the Cognitive Task differed can be seen in Figs (1-3). As seen in Fig. (1), Truthful, compared to Deceptive claimants had more to say

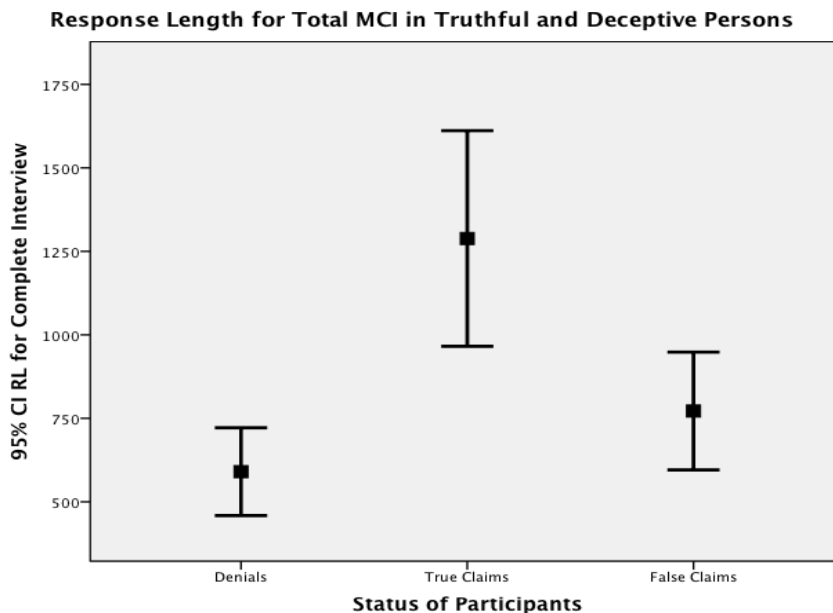


Fig. (1). Manual task total interview response length in MCI.

overall in the total interview. As seen in Fig. (2), however, when given the 1st prompt of the MCI (i.e. the details prompt) the Response Length of False Claims and Genuine claims were indistinguishable whereas the Response Length of denials was significantly diminished. As noted in Fig. (3), True claims could be distinguished from False claims by the relative lack of responsiveness to the mnemonic prompts by False Claims group. Similar patterns were also seen for the Cognitive Task.

In order to assess how well forensic statement data could be used to detect truthful from deceptive accounts, Binary Logistic Regression analyses (using Response Length from the Detail Prompt and the Response Length of the summed mnemonic prompts of the MCI for the Manual Task) were conducted. The model was significant ($p < 0.001$) and resulted in a classification accuracy of 80%. Similar analyses with

respect to the Cognitive Task indicated that the model was significant ($p < 0.008$) and resulted in a classification accuracy of 78%. This said, in the real world, professionals must make judgments about individual people or claims. Simply knowing that groups differ with respect to certain speech content variables may not significantly facilitate “N” of one decisions.

ROC calculations offer professionals a way to analyze data so as to make “N” of one decisions, rather than basing decisions on group norms. Consistent with the above, and noted in Tables 5 and 6, the ROC analyses provided evidence that a number of the speech content variables (for both the Manual and the Cognitive task) could be useful when trying to distinguish truthful from deceptive accounts. However for both tasks, the variable Prompt Total appeared to generate the most area under the curve.

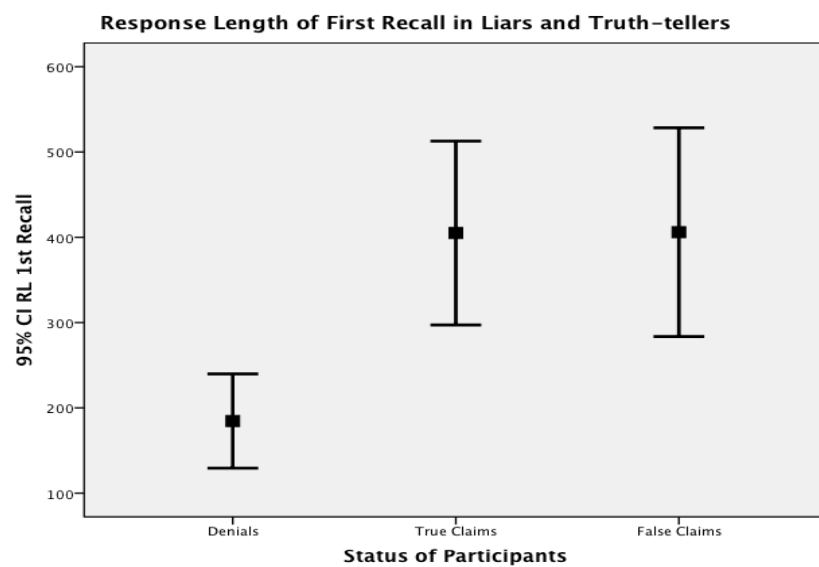


Fig. (2). Manual task response length in response to 1st MCI prompt (Details Prompt).

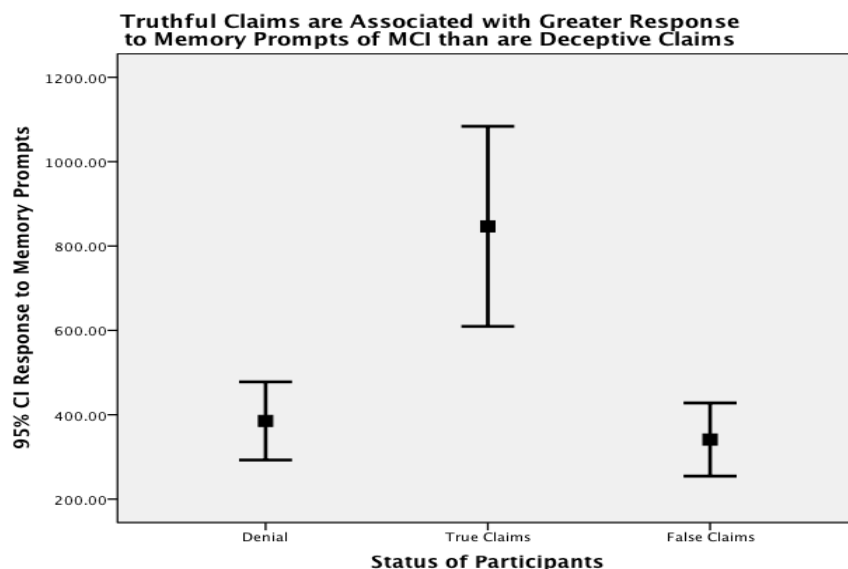


Fig. (3). Manual task response length to the mnemonic prompts (i.e. the Prompt Total) of the MCI.

Table 5. ROC data for sensitivity and 1-specificity (Manual Task).

| Table Result Variables(s) | Positive if Greater than or Equal to | Sensitivity | I-Specificity |
|---------------------------|--------------------------------------|-------------|---------------|
| Prompt Total | | | |
| | 75.00 | 1.00 | 1.00 |
| | 85.00 | 1.00 | 0.94 |
| | 108.00 | 1.00 | 0.89 |
| | 125.00 | 1.00 | 0.83 |
| | 162.00 | 0.94 | 0.83 |
| | 206.50 | 0.94 | 0.78 |
| | 232.50 | 0.94 | 0.72 |
| | 248.50 | 0.94 | 0.67 |
| | 255.50 | 0.94 | 0.61 |
| | 264.50 | 0.88 | 0.61 |
| | 275.50 | 0.88 | 0.56 |
| | 306.50 | 0.88 | 0.50 |
| | 368.50 | 0.88 | 0.44 |
| | 416.00 | 0.88 | 0.39 |
| | 435.50 | 0.88 | 0.33 |
| | 450.00 | 0.82 | 0.33 |
| | 461.00 | 0.82 | 0.27 |
| | 479.00 | 0.82 | 0.22 |
| | 498.50 | 0.77 | 0.17 |
| | 529.00 | 0.77 | 0.17 |
| | 551.50 | 0.77 | 0.11 |
| | 588.50 | 0.71 | 0.11 |
| | 627.00 | 0.71 | 0.06 |
| | 634.00 | 0.65 | 0.06 |
| | 649.00 | 0.65 | 0.000 |
| | 664.00 | 0.59 | 0.000 |
| | 684.50 | 0.53 | 0.000 |
| | 773.00 | 0.47 | 0.000 |
| | 906.50 | 0.41 | 0.000 |
| | 1017.50 | 0.35 | 0.000 |
| | 1069.00 | 0.29 | 0.000 |
| | 1107.00 | 0.24 | 0.000 |
| | 1169.00 | 0.18 | 0.000 |
| | 1468.00 | 0.12 | 0.000 |
| | 1775.00 | 0.06 | 0.000 |
| | 1809.00 | 0.00 | 0.000 |

*Includes data from all prompts

Table 6. ROC data for sensitivity and 1-specificity (Cognitive Task).

| Table Result Variables(s) | Positive if Greater than or Equal to | Sensitivity | I-Specificity |
|---------------------------|--------------------------------------|-------------|---------------|
| Prompt Total | | | |
| | 171.00 | 1.00 | 1.00 |
| | 200.50 | 0.95 | 1.00 |
| | 233.00 | 0.90 | 1.00 |
| | 238.50 | 0.90 | 0.86 |
| | 257.00 | 0.90 | 0.71 |
| | 289.50 | 0.85 | 0.71 |
| | 315.00 | 0.85 | 0.57 |
| | 347.50 | 0.80 | 0.57 |
| | 372.50 | 0.80 | 0.43 |
| | 404.50 | 0.80 | 0.29 |
| | 443.50 | 0.75 | 0.29 |
| | 470.00 | 0.70 | 0.29 |
| | 488.50 | 0.70 | 0.14 |
| | 504.00 | 0.65 | 0.14 |
| | 528.50 | 0.60 | 0.14 |
| | 544.50 | 0.55 | 0.14 |
| | 562.00 | 0.50 | 0.14 |
| | 656.00 | 0.50 | 0.000 |
| | 739.00 | 0.45 | 0.000 |
| | 772.00 | 0.40 | 0.000 |
| | 874.50 | 0.35 | 0.000 |
| | 1018.00 | 0.30 | 0.000 |
| | 1134.00 | 0.25 | 0.000 |
| | 1238.00 | 0.20 | 0.000 |
| | 1320.50 | 0.15 | 0.000 |
| | 1479.50 | 0.10 | 0.000 |
| | 1626.50 | 0.05 | 0.000 |
| | 1638.00 | 0.00 | 0.000 |

*Includes data from all prompts

Separate ROC calculations were performed using all significant variables from Tables 1 and 2; the Targeted Status was “truthful.” As shown in Tables 3 and 4, respectively, test results indicated that the Area under the Curve, and Asymptotic Significance for these variables was significant, indicating they could be used to sort claims. As noted in the tables, the most useful variable for both types of tasks was the *Prompt Total*.

DISCUSSION

Consistent with previous studies using cognitive interview methods for detecting deception, the present data indicate that, when exposed to cognitive interviewing techniques, genuine eyewitnesses have more to say and convey a greater level of detail in response to the mnemonic prompts. (Colwell *et al.*, 2002; 2007; Morgan *et al.*, 2007; 2008; 2008b; 2009; 2013; 2014). However, in addition to

what has previously been known about this issue, the present data indicate that when exposed to the MCI, genuine/truthful claimants, false claimant and denial participants behave in significantly different ways. These differences were seen when analyzing the transcripts and when assessing speech variables elicited by the MCI. Whereas genuine participants were the best distinguished from denial participants by their verbal responsiveness to the initial “detail” prompt of the MCI, genuine claimants were most effectively distinguished from false claimants by the verbal responsiveness of genuine persons to the second phase, (the mnemonic prompts phase) of the MCI. Unlike deniers who spoke 50% less than genuine claimants when given the initial MCI “details” prompt, false claimants, when exposed to the initial MCI “details” prompt, spoke a great deal and were indistinguishable from genuine claimants. However, when false claimants were subsequently exposed to the second phase of the MCI (i.e., the mnemonic prompts phase), false claimants spoke significantly less and were relatively non responsive to the memory recall techniques. In effect, after initially giving a detailed story, false claimants were unresponsive and offered little additional details about their memory. This finding has potential relevance to real world practices on the part of professionals who assess people for deception in that it suggests that false claimants may be indistinguishable initially from genuine claimants.

Within the context of this study, it is reasonable to hypothesize that the reduced responsiveness to the mnemonic prompts on the part of false claimants could be due to the increased cognitive load thought to be associated with a reasonably complex lie that entails their desire to “tell a plausible story” and stick to it so as to be believed. (Vrij et al, 2009). In essence, they have memorized a story they want to “sell” to the interviewer and they tell the entire “memory” when exposed to the initial prompt. As a result, these types of liars have little on which to elaborate when subsequently exposed to the mnemonic prompts.

The raters’ judgment data indicate that the raters performed significantly better than chance and performed significantly better than did professionals for whom accuracy judgments have been previously reported for experts tasked with rating videos of people exposed to the modified cognitive interview method (Morgan 2013; 2014). It is possible that they were better able to assess whether the claimant was truthful or lying because they read the transcripts and could visualize in a more concrete manner whether, and to what degree, participants had actually been responsive to the initial and mnemonic prompts of the MCI. Further studies may evaluate whether the accuracy can be increased using near real time speech to text translation.

The ROC data presented in this study demonstrate that if a database is collected, it is possible to apply the MCI derived forensic statement analysis methods to claims and to distinguish genuine from false claims about a task or activity in question. As noted in Tables 5 and 6, ROC data can be used as a way to make informed judgments about a person’s status (truthful/deceptive). For example, if one used *Prompt Total* data, the probability of being wrong (i.e. *1-Specificity*) in judging a Manual task claim containing 664 words (or more) as “genuine,” would be approximately 0.1%;

Similarly, the probability of being wrong in making this judgment if the claim had 588 or 461 words would be 11% and 28%, respectively. Similarly, if one used the cognitive task *Prompt Total* data when assessing a claim about the cognitive task, the probability of being wrong (i.e. *1-Specificity*) in judging a claim containing 656 words (or more) as “genuine” would be approximately 0.1%; Similarly, the probability of being wrong in making this judgment if the claim had 562 or had 470 words would be 14% and 29%, respectively.

If a professional wished to use actuarial speech content derived ROC data as a ‘laboratory test’ by which to determine the “truthfulness” of a claim, they could calculate a *likelihood ratio* known as *LR+*. *LR+* is defined as *Sensitivity/1-Specificity* and, unlike PPV and NPV tests, less affected by the actual prevalence of the targeted status (i.e. detecting truthful or deceptive claims) in the population. Calculating *LR+* is the equivalent of calculating the probability that a claim that “tests positive” truly is in the condition one is looking for (in this case, a genuine claim), divided by the probability that a claim that “tests positive” but is not in the condition we are looking for (i.e. not a genuine claim). If the *LR+* is greater than 1 this would indicate that the test result is associated with the “condition” we are looking for; if the *LR+* is less than one, this is associated with the absence of the “condition” we are looking for. However, *LR+* values that lie close to “1” have little practical import in that the “post test probability” is little different from the “pre-test probability.” As such, these scores might useful for diagnostic purposes but not for screening purposes. When *LR+* values are greater than or equal to 5 (or the *LR-* less than 0.2) they can be applied to the “pre-test” probability that the person (or claim) has the condition one is looking for (i.e. is genuine) – and serve well as a screening tool.

Accordingly, in this study, if, for the Manual task, we select a *Prompt Total* value of 306 words, the *LR+* would be: $.88/.50 = 1.76$. This value is very close to 1 and is unlikely to be useful in “screening” a set of “claims” to for identifying which are genuine claims. However, if we were to select a value of 627 words, the *LR+* would be $.71/.06 = 12$. This value would be very useful in pre-screening claims. Indeed, using the ROC table data in this study, an *LR+* of 5 or greater can be achieved by using a cutoff value of 529 words. Thus, if one were to screen a set of claims about the Manual task, and use this cutoff value, one would retain as “genuine” or “truthful” participants whose accounts had *Prompt Total* values of 529 words or more. The probability of being wrong in judging a person as genuine or truthful using this cutoff value is 11%. Using a similar approach for evaluating Cognitive claims, based on the data in Table 7, one would achieve an *LR+* value of 5 by using a *Prompt Total* cut off value of 488 words. The probability of wrongly classifying a truthful person using this criterion would be 14%.

This type of method could be employed in a number of settings in which professionals are tasked with making such judgments if a systematic collection of data is performed in order to form a data-base of interest regarding a targeted activities. Although some might object to creating a “data-

base” and cite, as the basis for the objection, the idea that it is impossible to anticipate all potentially relevant scenarios of interest, this objection, in our view, is misplaced. Within a particular discipline, there are high frequency issues of interest as well as “low frequency-high impact” issues of interest. We here suggest that the benefit of creating data-bases of interest is that they permit ROC based analyses that would serve LEO and intelligence professionals far better than any currently available vetting technologies.

Still, without creating reference data bases, the present data indicate that if one used the MCI technique, it is possible to make significantly superior detecting deception judgments than are afforded by currently marketed technologies such as the polygraph. Consistent with previous studies, the accuracy rates using MCI are well above the detecting deception accuracy rates documented for the polygraph (i.e. 56-60%) and demonstrated that cognitive load based models for detecting deception represent a valid and viable method for deception detection. In addition, The fact that the MCI method was effective for different types of activities (manual/cognitive) suggests that MCI may have a wider applicability – such in medical legal or insurance contexts in which professionals must often decide between true and false claimants. Future studies will examine whether the application of MCI within forensic and insurance contexts confers an advantage to adjudicators.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

- Altham, P.M.E. (1973). A non-parametric measure of signal discriminability. *British Journal of Mathematical and Statistical Psychology*, 26, 1-12.
- Bond, C.F., Kahler, K.N., & Paolicelli, L.M. (1985). The miscommunication of deception: an adaptive perspective. *Journal of Experimental Social Psychology*, 21, 331-345.
- Colwell, K., Hiscock, C.K., & Memon, A. (2002). Interview techniques and the assessment of statement credibility. *Applied Cognitive Psychology*, 16(3), 287-300.
- Colwell, K., Hiscock-Anisman, C., Memon, A., Rachel, A., & Colwell, L. (2007). Vividness and spontaneity of statement detail characteristics as predictors of witness credibility. *American Journal of Forensic Psychology*, 25(1), 5-30.
- DePaulo, B.M., & Pfeifer, R.L. (1986). On the job experience and skill at detecting deception. *Journal of Applied Social Psychology*, 16, 249-267.
- Ekman P., & O’Sullivan, M. (1991). Who can catch a liar? *American Psychologist*. 46 (9), 913- 920.
- Government Report (2009). *Detecting deception: Guidance on tools and techniques*. Center for the Protection of National Infrastructure (CPNI), British Security Service, Technical Report.
- Harnsberger, J.D., Hollien, H., Martin, C.A., & Hollien, K.A. (2009). Stress and deception in speech: Evaluation of layered voice analysis. *Journal of Forensic Sciences*, 54(3), 642-650.
- Hollien, H., Harnsberger, J.D., Martin, C.A., & Hollien, K.A. (2008). Evaluation of the NITV CVSA. *Journal of Forensic Science*, 53(1), 183-193.
- Metz, C.E. (1978). Basic principles of ROC analysis. *Seminars in Nuclear Medicine*, 8, 283-298.
- Morgan, C.A. III, Hazlett, G., Doran, A., Garrett, S., Hoyt, G., Thomas, P., Baranoski, M., & Southwick, S. (2004). Accuracy of eyewitness memory for persons encountered during exposure to highly intense stress. *International Journal of Law and Psychiatry*, 27, 265-279.
- Morgan, C.A. III, Steffian, G., Colwell, K., Coric, V., & Hazlett, G. (2007). Efficacy of forensic statement analysis in distinguishing true from false eyewitness accounts. *Journal of Intelligence Community Research and Development (JICRD)*, May, 1-35.
- Morgan, C.A. III, Mishara, A., Christian, J., & Hazlett, G. (2008). Detecting deception through automated analysis of translated speech: Credibility assessments of Arabic-speaking interviewees. *Journal of Intelligence Community Research and Development*, 1-22.
- Morgan III CA, Colwell K, Steffian G, Hazlett G. (2008b) Efficacy of verbal and global judgments in the detection of deception in moroccans interviewed via an interpreter. *Journal for Intelligence Community Research and Development*, 1-32.
- Morgan, C.A. III, Rabinowitz, R., Christian, J., & Hazlett, G. (2009). Detecting deception in vietnamese: efficacy of forensic statement analysis when interviewing via an interpreter. *Journal of Intelligence Community Research and Development*, 1-38.
- Morgan, C.A. III, Bhatt, S. (2012). A critical assessment guide to the topic of detecting deception: identifying methods that work and those that don't. *Journal of Intelligence Community Research and Development*, 1-76.
- Morgan, CA, Rabinowitz YG, Hilts D, Weller CE, Coric V. (2013). Efficacy of modified cognitive interviewing compared to human judgments in detecting deception related to bioterror activities. *Journal of Strategic Security*, 6 (3), 100-119.
- Morgan, III C.A., Rabinowitz ,Y.G., Leidy, R., Coric, V. (2014). Efficacy of combining interview techniques in detecting deception related to bio-threat. *Behav Sci Law*, 32, 269-285.
- Newman, M.L., Pennebaker, J.W., Berry, D.S., & Richards, J.M. (2005). Lying words: predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, 29(5), 665-675.
- SPSS (2010). SPSS Version 19. Armonk, NY: IBM
- Vrij, A., & Akehurst, L. (1998). Verbal communication and credibility: Statement validity assessment. In: Memon, A., Vrij, A., Bull, R. (Eds.), *Psychology and law: Truthfulness, accuracy and credibility*, pp 3-31 McGraw-Hill: Maidenhead
- Vrij, A. (2008). *Detecting lies and deceit: Pitfalls (Chapter 5)*, Chichester: Wiley pp. 373-388.
- Vrij, A., Fisher, R., Mann, S. & Leal, S. (2009). *Increasing cognitive load in interviews to detect deceit*. In: Williamson, T., Milne, B., & Savage, S., (eds). *International developments in investigative interviewing*, pp. 176-189 Willan Publishing: Uffculme.