

2010-01-01

# Intelligent Tutoring For Interviewing To Detect Deception: Can Investigators Be Trained To Attain And Detect Accurate Cues To Deception?

Justin Albrechtsen

*University of Texas at El Paso*, [jsalbrechtsen@miners.utep.edu](mailto:jsalbrechtsen@miners.utep.edu)

Follow this and additional works at: [https://digitalcommons.utep.edu/open\\_etd](https://digitalcommons.utep.edu/open_etd)



Part of the [Cognitive Psychology Commons](#)

---

## Recommended Citation

Albrechtsen, Justin, "Intelligent Tutoring For Interviewing To Detect Deception: Can Investigators Be Trained To Attain And Detect Accurate Cues To Deception?" (2010). *Open Access Theses & Dissertations*. 2628.  
[https://digitalcommons.utep.edu/open\\_etd/2628](https://digitalcommons.utep.edu/open_etd/2628)

This is brought to you for free and open access by DigitalCommons@UTEP. It has been accepted for inclusion in Open Access Theses & Dissertations by an authorized administrator of DigitalCommons@UTEP. For more information, please contact [lweber@utep.edu](mailto:lweber@utep.edu).

INTELLIGENT TUTORING FOR INTERVIEWING TO DETECT DECEPTION:  
CAN INVESTIGATOR BE TRAINED TO ELICIT AND DETECT ACCURATE  
CUES TO DECEPTION?

JUSTIN SCOTT ALBRECHTSEN

Department of Psychology

APPROVED :

---

Christian A. Meissner, Ph.D., Chair

---

Stephen L. Crites, Jr., Ph.D.

---

Harmon M. Hosch, Ph.D.

---

Matthew H. Scullin, Ph.D.

---

Samuel C. Riccillo , Ph.D.

---

Patricia D. Witherspoon, Ph.D.  
Dean of the Graduate School

Copyright © 2010  
Justin Scott Albrechtsen  
All Rights Reserved

To those who helped along the way

INTELLIGENT TUTORING FOR INTERVIEWING TO DETECT DECEPTION:  
CAN INVESTIGATOR BE TRAINED TO ELICIT AND DETECT ACCURATE  
CUES TO DECEPTION?

By

JUSTIN SCOTT ALBRECHTSEN

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

Department of Psychology

THE UNIVERSITY OF TEXAS AT EL PASO

May 2010

## ACKNOWLEDGEMENTS

I would like to express my gratitude to several people who have supported me in completing my degree. First, thank you to Llubia Corella who supported me on a daily basis. Second, thank you to my friends and family who encouraged and helped me through my time at UTEP. Third, thank you to my advisor, Dr. Christian Meissner and the Psychology Department.

## ABSTRACT

The present study examined whether investigators can be trained to elicit and detect cues to deception. The study included two training conditions and a control condition. Participants in the virtual Human Intelligent Tutoring System (vHITS) conditions completed a training program for deception detection and investigative interviewing. The primary components of this training were one-on-one interaction with a virtual human and tutoring tailored to specific participant responses. Participants in the Computer Based Training (CBT) conditions completed a comparable training program for deception detection and investigative interviewing. However, this program provided a more passive training environment with no interaction between student and tutor. Participants in the control conditions received no training. The performance of all three conditions on a deception detection task and an investigative interviewing task were compared. Results showed significant within-group differences between pre-test and post-test performance for both training conditions on all measures. However, results did not show significant between-groups differences at post-test for the training conditions on any measure. Possible explanations of these results are discussed as well as implications for future research.

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	v
ABSTRACT.....	vi
TABLE OF CONTENTS.....	vii
LIST OF TABLES.....	x
LIST OF FIGURES.....	xi
LIST OF APPENDICES.....	xii
CHAPTER 1.	
INTRODUCTION.....	1
Human Deception Detection.....	2
Cues to Deception.....	4
Nonverbal Cues to Deception.....	5
Verbal Cues to Deception.....	7
Criterion-based content analysis.....	8
Reality monitoring.....	9
Training Deception Detection.....	11
Training Police to Detect Deception.....	12
Training for CBCA and RM.....	15
Training for Verbal vs. Non-verbal Cues to Deception.....	17
Interviewing to Detect Deception.....	18
Interviewing to Increase Cognitive Load.....	19
Information Gathering vs. Accusatorial Interview Approaches.....	21
Virtual Human / Intelligent Tutoring Approaches.....	24

The Development and Effectiveness of Intelligent Tutoring.....	24
The Process of Intelligent Tutoring.....	25
Virtual Human Tutoring.....	26
Overview of the Proposed Study.....	27
CHAPTER 2.	
METHOD.....	30
Participants.....	30
Design.....	30
Materials.....	30
Pre-test and post-test stimuli.....	30
vHITS training program.....	33
CBT training program.....	37
No-training control filler task.....	39
Procedure.....	40
Phase 1: Pre-testing.....	40
Phase 2: Training on cues to deception and interviewing approaches.....	43
Phase 3: Post-testing.....	43
CHAPTER 3.	
RESULTS.....	44
Deception Detection Performance.....	44
Interviewing Performance.....	48
CHAPTER 4	
DISCUSSION.....	51

Deception Detection Performance.....	51
Interviewing Performance.....	54
Format and Length of Training.....	55
Issues with the vHITS Training.....	56
Potential for Application.....	59
General Discussion.....	60
REFERENCES.....	62
FOOTNOTES.....	71
CURRICULUM VITA.....	84

## LIST OF TABLES

Table 1	Mean and standard deviation estimates for performance on the deception detection task across training conditions and pre-test/post-test.....	72
Table 2	Mean and standard deviation estimates for performance on the knowledge for cues to deception and interviewing skills questionnaire across training conditions and pre-test/post-test.....	73

## LIST OF FIGURES

Figure 1	Mean discrimination accuracy ( $A_z$ ) scores at both pre-test and post-test across all conditions.....	74
Figure 2	Mean response bias (C) scores at both pre-test and post-test across all conditions.....	75
Figure 3	Mean confidence scores at both pre-test and post-test across all conditions.....	76
Figure 4	Mean scores for knowledge of diagnostic verbal cues to deception at both pre-test and post-test across all conditions.....	77
Figure 5	Mean scores for knowledge of non-diagnostic nonverbal behaviors at both pre-test and post-test across all conditions.....	78
Figure 6	Mean scores for knowledge of investigative interviewing techniques at both pre-test and post-test across all conditions.....	79

## LIST OF APPENDICES

Appendix A. Questionnaire for knowledge of cues to deception and interviewing skills.....80

## INTRODUCTION

How does one detect the difference between the truth and a lie? Deception is most likely very common within the legal system. Criminals may give false alibis about their whereabouts at the time of a crime (Olson & Wells, 2004), and innocent suspects sometimes provide false alibis and confessions (Leo & Ofshe, 1998). Considering these scenarios, the ability to detect deception would be a valuable skill for law enforcement. Unfortunately, research shows that lie detection is extremely difficult (DePaulo & Morris, 2004; Vrij, 2000; Vrij & Mann, 2003; Zuckerman, DePaulo, & Rosenthal, 1981), and even though law enforcement officials routinely encounter deception, they are not particularly adept at detecting it (Ekman & O'Sullivan, 1991; Kassin, Meissner, & Norwick, 2005). As will be discussed below, the inability of law enforcement officials to detect deception may be partially explained by the ineffective accusatorial interviewing tactics they use which fail to elicit accurate cues to deception from the interviewee (Vrij, Mann, & Fisher, 2006a; Vrij, Mann, & Fisher, 2006b).

Research on training for deception detection has employed a wide variety of methodologies and achieved mixed results. Many studies suggest that training is helpful if it focuses on indirect cues to deception (e.g., signs of tenseness or having to think hard) (DePaulo, Rosenthal, & Rosenkrantz, 1982; Mann & Vrij, 2006; Vrij, Edward, & Bull, 2001), while several other studies suggest that the training law enforcement officials generally receive does not increase accuracy, but rather creates over-confidence (Kassin & Fong, 1999; Meissner & Kassin, 2002). One meta-analysis suggests that training results in a small, but significant, improvement in deception detection performance even though the methodology used in most training studies is in many ways flawed (Frank & Feeley, 2003).

The present study assessed two training programs that were designed to improve deception detection and investigative interviewing capabilities by focusing participants on the most diagnostic (empirically validated) cues to deception. Literature in three areas will be reviewed. First, the human deception detection literature will be discussed with regard to (a) verbal and non-verbal cues to deception, (b) interviewing techniques that have been shown to accurately detect deception, and (c) research on training to detect deception. Second, research on interviewing to detect deception will be covered, including recent research on the merits of approaches that increase cognitive load and the relative benefits of information-gathering interviewing vs. accusatory interrogation approaches. Third, aspects of virtual human/intelligent tutoring approaches will be reviewed, including the effectiveness of these approaches, an explanation of the process of intelligent tutoring, and reasons why virtual human tutoring is beneficial when compared with more passive learning environments. The training programs and methodology used in the current study will also be explained, as will the measures and analyses used to assess the effectiveness of the proposed training programs.

### **Human Deception Detection**

For decades, researchers have explored how well humans detect deception. Bond and DePaulo (2006) conducted a meta-analysis on the accuracy of deception detection judgments. Across a sample of 206 studies, Bond and DePaulo found that individuals achieve an average of 54% accuracy for lie-truth judgments, by correctly identifying lies with 47% accuracy and correctly identifying truths with 61% accuracy. The results from within study analyses showed that lies are easier to discriminate from truths when the liar is motivated, unprepared, and has had prior exposure to the receiver of the lie. Results also showed no evidence that experts (e.g., law enforcement, judges, psychiatrists, etc.) are superior to non-experts in discriminating lies from

truths. Experts were more likely to have a deception bias such that they were more likely to judge statements as deceptive, whereas lay individuals generally demonstrated a truth bias such that they were more likely to judge statements as truthful (see also, Meissner & Kassin, 2002).

DePaulo et al. (2003) recently conducted a meta-analysis of the various verbal and non-verbal cues believed to distinguish truth vs. deception. The study included 120 independent samples and assessed 158 verbal and behavioral cues to deception. Results showed that the vast majority of the cues to deception were unreliable indicators, including number of blinks, posture, speaking rate, hand movements, number of pauses, and eye contact. However, their analysis did identify several other cues that proved to be reliable indicators of deception, including the number of negative statements, the amount of complaining, the number of ordinary imperfections, the overall sensibility of the story, admitted lack of memory, and spontaneous corrections.

Overall, research indicates that both law enforcement officials and lay individuals detect deception at about chance levels, though these populations vary in the extent to which they may be biased to perceive truth vs. deception. Moreover, many cues to deception that individuals believe to be diagnostic (and regularly touted by law enforcement training programs as such) have found little support in the empirical literature. The current study sought to identify the most diagnostic cues to deception by relying upon the empirical literature, and to thereby develop a training program that teaches participants to correctly identify the presence (or absence) of these cues in the context of an investigative interview. Furthermore, participants were trained to elicit these cues via appropriate interview techniques. The following sections detail the research literature supporting this approach.

## **Cues to Deception**

Explanations for why humans are unable to detect deception can be found by examining cues to deception. The popular story of Pinocchio tells of a little boy whose nose grew magically whenever he lied. Unfortunately in real life there is no growing nose or specific cue that makes lying immediately apparent to all (DePaulo & Morris, 2004). For years, both researchers and practitioners have been seeking to identify the most diagnostic cues to deception.

### **Beliefs about Cues to Deception**

Understanding people's beliefs regarding cues to deception is an important first step in explaining their deception detection capabilities (Vrij, 2008). Forrest, Feldman, and Tyler (2004) tested the hypothesis that accurate beliefs about cues to deception would be correlated with accuracy on a deception detection task. To test this hypothesis, participants completed an 18-item belief questionnaire on cues to deception (DePaulo, Stone, & Lassiter, 1985) after performing a deception detection task. Examples of accurate beliefs on the questionnaire include beliefs that false statements contain more negative statements and irrelevant information than true statements. Examples of inaccurate beliefs include beliefs that eye contact, smiling, and shrugging are indicative of deception. Results confirmed the hypothesis that participants with accurate beliefs about cues to deception would outperform participants with inaccurate beliefs about cues to deception on a deception detection task.

By surveying people regarding their beliefs about cues to deception and comparing the results to the frequency of actual coded behaviors in deceptive statements, researchers can assess how correct these beliefs are. For example, Akehurst, Köhnken, Vrij, and Bull (1996) examined 60 lay persons' and 60 police officers' beliefs about deceptive behavior. Participants completed a 64-item questionnaire on beliefs regarding deceptive behavior. The questionnaire covered 4

categories of cues: facial behaviors, bodily behaviors, speech characteristics, and content of the statement. Results showed that both police officers and lay persons tended to believe that changes in eye contact, shifts in posture, blinking, fidgeting, and smiling are indications of lying. In addition, both groups believed that increased hand, arm, foot, and leg movements were also cues to deception. The results of this study also showed that beliefs about cues to deception differ very little between laypersons and law enforcement professionals, with both groups purporting that the most accurate cues to deception are non-verbal cues. Unfortunately, people's beliefs about cues to deception are generally incorrect – as discussed below, the most diagnostic cues tend to involve aspects of the verbal content or speech acts of the sender.

### **Nonverbal Cues to Deception**

Non-verbal cues to deception include any aspect of a liar's behavior that can be identified without the use of sound (Burgoon, Blair, & Strom, 2008). Some examples of non-verbal cues include gaze aversion (e.g., not making eye contact), posture (e.g., crossing one's arms), facial expressions (e.g., smiling, frowning, or scowling), tics (e.g., rubbing one's head/neck/chin), gestures (e.g., pointing at someone), hand and arm movements (e.g., making a circular motion with one's arms), and head movements (e.g., shaking/nodding one's head). Research shows that non-verbal cues are much less diagnostic than verbal cues (Bond & DePaulo, 2006; Vrij, 2008). As described above, DePaulo et al. (2003) conducted a meta-analysis of 158 cues to deception, finding significant differences between liars and truth-tellers for only a small number of non-verbal cues included in the study. Importantly, no significant differences were found between liars and truth-tellers for many of the most commonly believed non-verbal cues to deception, including eye contact, blinking, fidgeting, smiles, and postural adjustments.

Sporer and Schwandt (2007) conducted a meta-analysis of non-verbal cues to deception and assessed the role of several moderator variables across studies. Of the eleven different non-verbal behaviors analyzed, only three (nodding, foot and leg movements, and hand movements) were reliably associated with deception. It is noteworthy that each of these three non-verbal behaviors was negatively correlated with deception, meaning that liars performed them less often than truth tellers. This is contrary to the commonly held belief that non-verbal behaviors increase when one is lying. Thus, even when a correlation exists between a non-verbal cue and deception it is usually in the opposite direction of what the commonly held belief would predict (Vrij, 2008).

One type of non-verbal cue that has received a great deal of attention is micro-expressions. A micro-expression is a brief, involuntary facial expression shown on the face of a human when one is trying to conceal an emotion (Russell, Chu, & Phillips, 2006). Tics, furrows, smirks, frowns, smiles, and wrinkles emerge in assorted combinations that are outside of an individual's control. For example, one might raise an eyebrow slightly when surprised or barely close the eyelids to squint when concentrating. Some experts believe that micro-expressions betray a liar's true emotion and are therefore a useful tool in deception detection (O'Sullivan & Ekman, 2004).

A program within the Department of Homeland Security called Screening of Passengers by Observation Techniques (SPOT) is now training Transportation Security Administration (TSA) agents to use micro-expressions and other non-verbal behaviors to identify suspicious passengers. A passenger exhibiting these behaviors is flagged and must undergo additional screening. The SPOT program has already been deployed in several airports, and a nationwide rollout is planned for the near future

(see [http://www.tsa.gov/press/where\\_we\\_stand/training.shtm](http://www.tsa.gov/press/where_we_stand/training.shtm)).

Porter and ten Brinke (2008) recently provided one of the first independent assessments of the presence of micro-expressions in truthful and deceptive individuals. Results showed that participants performed only slightly above chance levels at determining which facial expressions were truthful vs. deceptive. In addition, micro-expressions were exhibited by only 22% of participants in approximately 2% of all expressions, and in the upper or lower face only, suggesting that even if micro-expressions are accurate cues to deception they are not expressed often enough by liars to really be helpful.

### **Verbal Cues to Deception**

While non-verbal cues to deception are typically non-diagnostic, many verbal cues accurately distinguish between lies and truths (DePaulo et al., 2003). A verbal cue to deception is any behavior that can be identified just by reading the liar's statement (Vrij, Edward, Roberts, & Bull, 2000). Some examples of verbal cues include the structure of a statement (e.g., the chronological ordering of events), spontaneous corrections (e.g., identifying an incorrect detail), the amount of detail in a statement (e.g., providing sensory information), and negative statements (e.g., complaining about questions).

In the previously mentioned meta-analysis of cues to deception, DePaulo et al. (2003) identified 23 cues with large effect sizes for significant differences between liars and truth tellers. Of these 23 cues, 15 were strictly verbal cues and another 3 were partial verbal cues. Some of these cues are not surprising. For example, liars' statements contain less logical structure than the statements of truth tellers. However, other cues are somewhat counterintuitive. For example, liars' statements actually contain fewer spontaneous corrections and instances of admitted lack of memory than do the statements of truth tellers. Most of these cues are

performed less by liars than truth tellers. For example, liars provide fewer details, especially temporal and spatial details, than the statements of truth tellers. Still, a few of these cues occur more often in liars' statements than in truth tellers' statements. For example, liars' statements contain more negative statements and complaints than the statements of truth tellers. Because these verbal cues are among the most reliable and accurate of all cues to deception, they can be found in a wide variety of effective deception detection techniques.

One method to examine the diagnosticity of verbal vs. non-verbal cues is to manipulate the medium used to present deceptive or truthful stimuli. For example, if a participant were shown only video, he/she will be able to perceive both verbal and non-verbal cues, whereas presenting only audio with no video allows the participant to perceive only verbal cues. Kassin, Meissner, and Norwick (2005) conducted a study in which college students and police investigators watched or listened to 10 confessions (5 true, 5 false) of prison inmates. Results showed that accuracy rates were significantly greater for both students and police when judgments were made using audiotape rather than videotape. Bond and DePaulo's (2006) meta-analysis of deception detection studies also confirmed this finding – accuracy rates were significantly higher for audio only conditions (63% accuracy) than for video only conditions (52% accuracy) or audio-visual conditions (56% accuracy), suggesting that verbal cues to deception are more diagnostic than non-verbal cues.

**Criteria-based content analysis.** One successful deception detection technique which focuses on verbal cues is Criteria-Based Content Analysis (CBCA). CBCA is the third stage of a four stage process known as Statement Validity Assessment (SVA). SVA was initially developed to determine the truthfulness of child witnesses/victims in sexual abuse cases, but the technique is now used in many deception detection contexts involving adults (Vrij, 2008).

CBCA involves a systematic assessment of the credibility of a written statement. By assessing only the written statement, CBCA only takes into account verbal cues to deception, while ignoring non-verbal cues to deception (Vrij, Akehurst, Soukara, & Bull, 2004). CBCA evaluates statements based on the presence or absence of 19 criteria. Many of these criteria involve the same reliable verbal cues found in the Depaulo et al. (2006) meta-analysis of cues to deception, including amount of detail, temporal/spatial detail, logical structure, negative statements, admitted lack of memory, and spontaneous corrections.

Vrij (2005) conducted a qualitative review of the first 37 studies using CBCA. In 11 of the 12 laboratory studies included in the analysis, CBCA scores for truthful statements were significantly higher than for false statements. In field studies, CBCA scores on average were higher for truthful statements than they were for false statements, although this effect was not as robust as the laboratory studies. Results also showed that inter-rater reliability for each of the 19 CBCA criteria was good (.60 to .75) to excellent (.75 or above) and inter-rater reliability for the overall CBCA score was excellent (however, see Ruby & Brigham, 1997). Based on these results, Vrij argued that the CBCA could be a useful tool for conducting police investigations.

**Reality monitoring.** Another technique using only verbal cues to detect deception is Reality Monitoring (RM). RM is based on the belief that true memories of real events are qualitatively different from false memories of imagined events (Sporer, 2004). Memories of real events involve the intake of environmental information through the sensation and perception of this information by the brain. Therefore, these memories tend to contain more sensory (e.g., details of taste, smell, touch, sound, and vision), contextual (e.g., details of where and when the event happened), and affective information (e.g., details of feelings and mood; see Johnson & Raye, 1998). In contrast, false memories of imagined events are created rather than remembered,

and therefore do not involve the processes of sensation and perception. False memories thereby lack sensory, contextual, and affective information and instead contain evidence of cognitive operations (e.g., thoughts and reasons). That is, when people create stories they tend to make their stories interesting and easy to understand. When memory fails, gaps are bridged by including information that seems believable but isn't actually remembered (Johnson, 1988).

The process of reality monitoring is similar to that of CBCA (Vrij, 2008). A written statement is analyzed for the presence of various criteria. Unlike CBCA, however, there is no standardized set of criteria for RM. Instead each researcher uses his/her own set of RM criteria, which may be similar to many of the CBCA criteria and verbal cues found to be reliable in the DePaulo et al. (2006) meta-analysis of cues to deception. For example, amount of detail, temporal/spatial detail, logical structure, negative statements, admitted lack of memory, and spontaneous corrections are criteria that are frequently used in RM studies.

Sporer (2004) conducted a quantitative review of RM studies. Because each researcher uses his/her own RM criteria, it is difficult to compare results across studies. Many studies showed significant differences between RM scores for deceptive vs. truthful statements, although some studies showed only small differences in RM scores. There are many possible explanations for this variation across studies, as such effects may be moderated by factors such as the opportunity of the liar to prepare his/her lie, the age of the liar, and the scoring system used. Several of the studies that failed to find significant differences included RM criteria that were not derived from deception detection research. For example, in some RM studies confidence in one's memory was considered an indicator of truth – however, the deception detection literature has shown that admitting a lack of memory is an indicator of truth. Sporer

concluded that the best results in deception detection accuracy would likely come from combining common aspects of RM and CBCA.

### **Training Deception Detection**

Improving deception detection capabilities is a difficult undertaking. Research on training for deception detection has employed a wide variety of methodologies and achieved mixed results. Not surprisingly, training studies that focus on reliable verbal cues to deception are most successful in achieving high accuracy rates.

Frank and Feeley (2003) conducted a meta-analysis on the effectiveness of training deception detection. In a sample of eleven studies, a mean effect size of  $r = .20$  was found across 20 paired comparisons of lie detection training conditions versus no-training control conditions, suggesting that training does significantly raise lie detection accuracy rates. Frank and Feeley proposed that six criteria were necessary to develop effective training studies in this area. Criteria 1 and 2 state that training studies must include relevant real-world interrogation settings and high stakes lies. These criteria allow the findings in training studies to generalize to more applied settings. Criteria 3 and 4 state that training studies must include proper training as well as proper pre-testing and post-testing of deception detection performance. These criteria are necessary to establish base-line performance, improve this performance, and then establish whether or not performance has changed. Finally, criteria 5 and 6 state that the results of training studies should be generalizable over time and across deception detection situations. By including these criteria, training will be useful and practical for real-world implementation. The authors note that most current training studies fall short of meeting these criteria, resulting in an underestimation of the positive effects of deception detection training.

## **Training Police to Detect Deception**

Research shows no evidence that law enforcement officials are superior to non-experts in discriminating lies from truths (Bond & DePaulo, 2006). Meissner and Kassin (2002) examined the effects of experience and training on deception detection capabilities. They were particularly interested in whether law enforcement investigators might perform better at such a task when compared with lay (or untrained) individuals. Meissner and Kassin introduced the use of signal detection theory (SDT) to assess the effects of these variables on deception detection performance. SDT separates deception detection performance into two parameters, namely discrimination accuracy and response bias. Discrimination accuracy is generally defined as the ability of the individual to correctly detect a signal (e.g., deception) or reject its absence, while response bias is the degree of evidence (e.g., cues to deception) that an individual needs to identify a signal as being present and thereby assesses an individual's tendency to overly respond that a statement is either truthful or deceitful. Using SDT, Meissner and Kassin demonstrated that neither experience nor training improved discrimination accuracy (the ability of an individual to correctly distinguish between true and false statements). However, both experience and training did result in a "deception" response bias such that with more experience and training an individual was more likely to judge statements as deceptive. So while lay individuals are more likely to demonstrate a truth bias, investigative experience and training in deception detection are more likely to produce a bias towards perceiving "deception."

One popular technique used to train police officers to detect deception is the Reid Technique (Inbau, Reid, Buckley, & Jayne, 2001). This technique begins by encouraging the assessment of a suspect's veracity via the conduct of a Behavior Analysis Interview (BAI). The goal of the BAI is to ascertain whether a suspect is being truthful or lying. During the BAI, the

suspect is asked “behavior-provoking questions that are specifically designed to evoke behavioral responses” (Inbau et al., 2001). The Reid Technique expects liars to be less helpful than truth-tellers in investigations and to exhibit more non-verbal, nervous behaviors (e.g., fidgeting, hand and leg movement, etc.). Based on these behaviors, the interviewer decides whether the suspect is being truthful or lying. If the suspect is deemed a liar by the interviewer, an interrogation phase will begin. Reid proposes that officers are over 80% accurate in these judgments when they use the BAI.

While the Reid Technique (Inbau et al., 2001) states that liars are less helpful than truth-tellers in investigations and exhibit more nervous behaviors, previous deception detection literature predicts just the opposite (Vrij, 2008), suggesting that liars are actually more helpful and appear more relaxed than truth-tellers. Vrij, Mann, and Fisher (2006a) conducted the first empirical testing of the BAI. For this experiment, 40 participants lied or told the truth during a BAI interview. Following the interview their responses were coded according to the guidelines of the Reid Technique. Results were consistent with prior deception detection research. Specifically, truth tellers were less helpful and exhibited more nervous behaviors than liars.

Kassin and Fong (1999) examined the effects of training on judgments of truth and deception in the interrogation room. Forty participants were randomly assigned to two conditions. The participants in the experimental condition were trained using the Reid method of BAI, while those in the control condition received no training. All participants then evaluated videotaped denials of mock crimes, half of which were true and half of which were false. Results showed that trained participants were less accurate, but more confident, in their judgments than were non-trained participants, suggesting that law enforcement officials may be more confident in their deception detecting capabilities simply due to the fact they are receiving

training, even if the training is ineffective (a result subsequently confirmed by Meissner & Kassin, 2002).

No two liars deceive in exactly the same way (DePaulo & Morris, 2004). Vrij and Graham (1997) investigated whether simply informing police officers about individual differences between liars in deceptive behavior would facilitate better accuracy in a deception detection task. Forty college students and twenty nine police officers participated in the study. Participants viewed 10 videotaped statements of college students either lying or telling the truth. Half of the participants received information about the individual differences between liars in deceptive behavior while the other half did not. Results showed that college students were more accurate detectors of deception when receiving the information, while no significant differences were found between conditions for police officers.

Porter, McCabe, Woodworth, and Peace (2007) examined the influence of high motivation and accurate feedback on deception detection. Participants were randomly assigned to conditions according to motivation level (high or low) and feedback (accurate, inaccurate, or none). All participants gave veracity judgments for 12 videotaped statements (6 true, 6 false). Prior to completing this task, the experimenters manipulated motivation by telling half of the participants that a small number of people are excellent at detecting deception and that if they performed well in the experiment they could receive up to \$50. The other half of participants were told that lie detection is extremely difficult and they should simply try their best during the experiment. Following their veracity judgment for each statement, participants were accurately informed if their decision was correct/incorrect, were inaccurately informed if their decision was correct/incorrect, or received no feedback regarding their decision. While no main effects of feedback were found, it was surprising that accuracy rates were higher for the low-motivation

condition (60% accuracy) than for the highly motivated participants (46% accuracy) suggesting motivation is actually detrimental when attempting to detect deception. This finding does not bode well for police officers who are likely to be highly motivated to detect suspect's lies.

Research on the effects of motivation and training on law enforcement officials paints a bleak picture with regard to police officers' ability to detect deception. In general, popular training programs, such as those using the Reid Technique, train law enforcement officials to attend to unreliable, non-verbal cues to deception. As a result of this training, investigators fail to improve their accuracy, though they gain over-confidence in their performance and a bias towards viewing statements as deceptive. In addition, police officers are likely to be highly motivated to detect suspects' lies, though research suggests that motivation may actually be detrimental to deception detection performance. Unfortunately, while techniques like that of Reid are not based in empirical research on deception detection, they are still being used to train local, state, and federal law enforcement, as well as various human intelligence gatherers in the Department of Defense and intelligence community.

### **Training for CBCA and RM**

Empirically-derived alternatives to the Reid Technique, such as CBCA and RM, exist. Many studies have been conducted examining the effectiveness of training participants using these two techniques. The methodologies used in these studies vary. Some studies analyze only CBCA (Landry & Brigham, 1992) or only RM (Sporer & Sharman, 2006), while others compare CBCA and RM to each other (Sporer, 1997) or combine the two techniques (Vrij, Edward, Roberts, & Bull, 2000). One study examined the effectiveness of rapid judgments and shorter training for CBCA and RM (Vrij, Evans, Akehurst, & Mann, 2004), while another examined the differences between law enforcement and lay persons in CBCA training (Akehurst, Bull, Vrij, &

Köhnken, 2004). Even though the methodologies in these studies varies, their results are encouraging showing that by using the reliable verbal cues to deception common to these two techniques high accuracy rates can be achieved.

Many studies show the utility of CBCA and RM training. For example, in one study, the accuracy rate of CBCA trained participants on a deception detection task was both significantly better than chance and significantly better than a control condition (Landry & Brigham, 1992). In another study, researchers using the RM technique found that accounts of experienced events contained more qualitative details than accounts of imagined events (Sporer & Sharman, 2006). Vrij, Edward, Roberts, and Bull (2000) compared the accuracy rates of four deception detection techniques. Results showed overall accuracy rates of 73% for CBCA and 67% for RM. However, an overall accuracy rate of 81% was achieved when CBCA and RM were combined, suggesting that combining CBCA and RM can result in even greater accuracy rates than using either of these techniques individually.

One criticism of CBCA and RM is the amount of time needed to properly use these techniques. Each step of CBCA and RM is time-consuming. First, training can take hours or even days, and it generally requires an expert instructor. Next, statements must be transcribed, a monotonous process that at times is impractical for real world settings. Finally, the analysis of the transcripts is tedious and requires multiple raters (Vrij, 2008). Vrij, Evans, Akehurst, and Mann (2004) sought to make CBCA and RM less time consuming by shortening training, using videos of liars rather than written transcripts, and simplifying the actual process of CBCA and RM analysis. Five participants received 90 min of training from a CBCA and RM expert. Following training each of the five participants watched 52 video-taped interviews. For each of the videos, participants estimated the frequency of nine CBCA and RM criteria. Results showed

that participant ratings correlated with expert ratings and reliably discriminated between true and false statements. These results provide evidence that CBCA and RM can be trained and used quickly, cheaply, and effectively.

These studies show that people can be trained to use the CBCA and RM techniques to detect deception. By combining CBCA and RM, individuals can detect deception with greater accuracy than with either technique alone. While CBCA and RM are currently time-consuming processes, research suggests that CBCA and RM can be trained and used more quickly.

### **Training for Verbal vs. Non-verbal Cues to Deception**

Studies repeatedly show the effectiveness of training participants in the CBCA and RM techniques. The effectiveness of these techniques may be due to the fact that by using written transcripts of statements, only verbal cues to deception are available for analysis by the rater. As discussed previously, research consistently shows that verbal cues to deception are more reliable than non-verbal cues (DePaulo et al., 2003). This section will examine research comparing training focusing on verbal cues to deception to training focusing on non-verbal cues to deception to identify which type of training results in better deception detection accuracy.

Santarcangelo, Cribbie, and Ebesu Hubbard (2004) examined which of three types of training cues (visual, vocal, or verbal) would be most effective in improving deception detection accuracy. Results showed that accuracy rates were significantly better for participants trained on verbal cues. This overall improvement was largely due to participants' ability to detect truthful messages, as there were no significant differences between groups with regard to detecting deceptive messages. This result is encouraging in that it shows that training can improve participants' deception detection performance without sacrificing their tendency to believe that people are telling them the truth.

Sporer and Burch (2003) also examined the effectiveness of training participants to focus on verbal cues. Two hundred stories were rated by participants both before and after training. Training consisted of 21 verbal criteria such as logical consistency, information about space and time, and amount of detail. Prior to training, participants' deception detection performance was at about chance levels (50% accuracy). After training, deception detection performance improved significantly. Applying signal detection theory, it was shown that the advantage of training again did not come at the expense of a shift in response bias.

While the above studies show positive effects for training on verbal cues, other studies suggest that non-verbal cue training can enhance detection (deTurck, Feeley, & Roman, 1997; Feeley & deTurck, 1997), even when such cues are among those with the least diagnostic evidence (DePaulo et al., 2003). Still other studies show null effects when examining the effect of training for verbal vs. non-verbal cues to deception (DePaulo, Lassiter, & Stone, 1982; Köhnken, 1987).

Clearly training people to detect deception is not an easy task. Research shows that while police officers' training in unreliable non-verbal cues to deception does not generally result in improved deception detection performance, it does produce a deception bias, overconfidence, and difficulty grasping more effective deception training using reliable verbal cues. The training literature also demonstrates that training participants in techniques that focus on verbal rather than non-verbal cues, such as CBCA and RM, results in improved deception detection capabilities.

### **Interviewing to Detect Deception**

An important step in the process of accurately detecting deception is the elicitation of deceptive cues through interviewing. Research shows that some interviewing techniques, such

as those that increase cognitive load, elicit more cues to deception than others, and that increasing the amount of conversation via open-ended interview approaches can increase the possibility of eliciting verbal cues to deception. This section will also examine studies that have assessed investigative interviewing techniques, including information gathering approaches (e.g., the Cognitive Interview) and accusatorial approaches (e.g., Reid Technique), and their influence on deception detection.

### **Interviewing to Increase Cognitive Load**

Several studies show that lying is a more cognitively demanding task than telling the truth (DePaulo et al., 2003; Mann & Vrij, 2006; Vrij et. al., 2008; Zuckerman, DePaulo, & Rosenthal, 1981). Therefore attending to signs of cognitive load should increase deception detection accuracy. Cognitive load is a term that refers to the amount of duress placed on working memory when an individual is asked to perform various tasks (Barrett, Frederick, Haselton, & Kurzban, 2006).

Mann and Vrij (2006) conducted a study assessing whether signs of cognitive load would increase when one is lying. In this study 84 police officers watched 14 videos of actual suspects during their police interviews. Half of these suspects were lying and half were telling the truth. The control condition was asked after each video to indicate whether the suspect was lying or telling the truth. The experimental condition was asked after each video to identify in the suspects the extent of three manifestations of cognitive load: (1) tenseness, (2) having to think hard, and (3) attempting to control behavior. Results showed that by attending to signs of cognitive load, the experimental condition was more accurate at detecting deception than the control condition.

The majority of research focuses on training individuals to attend to reliable cues, and to ignore unreliable cues (Frank & Feeley, 2003; Vrij, Edward, & Bull, 2001). Vrij et al. (2008) created an alternative approach to deception detection which focuses on the liar himself rather than the individual attempting to detect the lie. This approach involves introducing a manipulation that increases the liar's cognitive load and thereby impairs his or her ability to tell a believable lie (Vrij, Fisher, Mann, & Leal, 2006).

Vrij et al. (2008) conducted a two-part study introducing a cognitive load manipulation which magnified the differences between lies and truths, making it easier to distinguish between the two. In Experiment 1, 80 mock suspects either lied or told the truth about a staged event during a videotaped interview. Participants in the control condition were asked to recall the events in their statements in chronological order. Participants in the experimental condition were asked to recall the events in reverse chronological order to increase cognitive load. Results showed that the statements given in reverse chronological order contained many more cues to deceit than the control interviews. In Experiment 2, 55 police officers watched these videotaped interviews and were asked to indicate whether each suspect was lying or telling the truth. Results showed that police officers were better able to detect deception when evaluating the reverse statements. This improvement in deception detection performance did not result in a response bias.

Vrij et al. (2008) showed that the interview used to elicit a statement can have an impact on how many cues to deception will be present in the statement. By simply asking participants to give their statement in reverse chronological order (thereby increasing cognitive load), cues to deception increased and deception was more easily detected. These results support the idea that

one of the goals of police interviews should be to elicit accurate cues to deception from the interviewee.

### **Information Gathering vs. Accusatorial Interview Approaches**

Police interviews are generally conducted with one of two goals in mind: gathering information or obtaining a confession. If the goal of the interview is to gather information, the interviewer will use open-ended questions in order to obtain detailed statements from the interviewee (e.g., “What did you do yesterday between 9 pm and 12 am?”). However, if the goal of the interview is to get a confession, the interviewer will often confront the suspect with direct accusations (e.g., “You robbed the gas station!”), resulting in abrupt denials from the accused (e.g., “I didn’t rob the gas station!”) (Fisher, Brennan, & McCauley, 2002; Vrij, Mann, & Fisher, 2006a; Vrij, Mann, Kristen, & Fisher, 2007). The information-gathering interview style is preferable to the accusatorial style for many reasons. Information-gathering interviews result in more information which police can compare to available evidence (Vrij, 2008). Statements gathered through the information-gathering interview style result in longer statements which contain more verbal cues to deception (Vrij et al., 2007). Accusatory interrogations are more likely to result in false confessions than information-gathering interviews (Gudjonsson, 2003). Finally, cognitive-based interventions (e.g., asking the interviewee to talk about the event in reverse chronological order) are easier to use in information-gathering interviews than in accusatory interrogations.

The cognitive interview (CI, Fisher & Geiselman, 1992) is one of the most well researched information gathering interview styles (Fisher & Schreiber, 2007; Köhnken, Milne, Memon, & Bull, 1999). The goal of the CI is to provide the interviewer with a systematic approach to elicit the maximum amount of relevant information from an interviewee. To achieve

this goal the CI employs techniques that have their basis in cognitive psychology, specifically research on retrieving information from memory (Vrij, 2008). Whereas traditional interviews attempt to *extract* information from memory by asking direct questions (e.g., How tall was the perpetrator?), the CI seeks to *elicit* information from memory by asking open-ended questions (e.g., What can you tell me about the perpetrator?). When conducted correctly the CI will consist of a few short open-ended questions and several long narrative responses by the interviewee. In order to do this, the interviewer must set an environment where the interviewee feels comfortable generating information without necessarily having to be prompted (Fisher & Geiselman, 1992). Stage two of Statement Validity Assessment (SVA) is a semi-structured interview where the witness provides his or her own account of the event (Vrij, 2008). Research shows that when this semi-structured interview is replaced by the CI, witness statements receive higher CBCA scores (Fisher, Brennan, & McCauley, 2002). The British policing system has moved toward the use of an information gathering approach, including aspects of the CI, with success for both witness and suspect interviews (Bull & Milne, 2004).

Unfortunately, many police officers are still trained primarily in accusatorial interrogation techniques such as the previously mentioned Reid Technique. As noted previously, the Reid Technique begins with the BAI, the goal of which is to ascertain whether a suspect is being truthful or lying. If the suspect is deemed a liar by the interviewer, an interrogation phase will begin. During this phase, the goal of the interview shifts from attempting to detect deceit to attempting to gain a confession from the suspect. In order to accomplish this task the interrogator will confront the suspect, shift blame away from the suspect, and never allow the suspect to deny guilt.

Vrij, Mann, and Fisher (2006b) examined interviewees' perception of both information gathering and accusatory interview styles. Forty participants were interviewed while either lying or telling the truth about an event. Interviewers used either an information gathering or accusatory interviewing style. Following the interview, participants rated their degree of discomfort and cognitive demand, as well as the degree to which they felt they had been listened to. Results showed that information-gathering interviews were perceived as more cognitively demanding, although accusatory interviews were perceived as making respondents more uncomfortable. Respondents felt that they were listened to more in information-gathering interviews. In addition, liars found the interviews to be more cognitively demanding than truth tellers.

Vrij, Mann, Kristen, and Fisher (2007) examined the effects of interview styles on the ability to detect deception. Sixty-eight police officers watched videotaped interviews and made veracity judgments. Interviews were conducted using an accusatory, information-gathering, or BAI approach. Results showed that watching accusatory interviews resulted in more false positives (accusing truth tellers of lying) and higher confidence than either the information-gathering or controlled interviews. No significant differences were found between groups on accuracy. These results suggest that accusatory interview approaches increase bias and confidence but not accuracy.

Taken together, the results of these studies show that investigative interviewing tactics are preferable to accusatory interrogation tactics for several reasons. Accusatory interrogations make interviewees uncomfortable and elicit signs of nervousness that are then misinterpreted by police as non-verbal cues to deception (e.g., fidgeting, hand and leg movement). However, investigative interviewing practices make the interviewee more comfortable resulting in

increased cooperation with the interviewer. Investigative interviewing also increases the cognitive load of the interviewee, thereby eliciting more reliable verbal cues to deception which make the liar more easily identifiable.

### **Virtual Human / Intelligent Tutoring Approaches**

One training technique that has not yet been examined within the context of detecting deception is the use of virtual human intelligent tutoring. An Intelligent Tutoring System (ITS) involves any computer system that provides direct customized instruction or feedback to students while performing a task without the help of human beings (Psotka, Massey, & Mutter, 1988). This section will explain why ITS appears to be effective, the process of an ITS, and why an ITS may be preferable to other training techniques.

### **The Development and Effectiveness of Intelligent Tutoring**

Anderson, Corbett, Koedinger, and Pelletier (1995) reviewed the 10 year history of intelligent tutoring development. Initially, computer tutors were based on cognitive models of how students solved problems in geometry and algebra. These ITS's were developed based on the principles created in Advanced Computer Tutor (ACT) theory (Anderson, 1983), a theory describing how skills are acquired via computers. Early evaluations of these tutors were promising. By using an ITS, students could achieve the same level of proficiency as conventional instruction, but ITS's only took one-third of the time that conventional instruction required.

Analyses of these early ITS's showed that the best results were attained when the tutor provided (a) detailed instructions to the student, (b) a problem-solving interface, and (c) immediate feedback to the student regarding his/her performance. Newer tutoring systems began to be developed which incorporated these factors into their designs. With these improvements,

ITS's became more user friendly and easily implemented in a wide variety of settings. The use of ITS's has now rapidly spread to many areas including technical military training, geography, circuits, medical diagnosis, computer programming, mathematics, physics, genetics, and chemistry (Koedinger & Corbett, 2006).

### **The Process of Intelligent Tutoring**

A typical ITS consists of four different modules: the interface module, the expert module, the student module, and the tutor module. The interface module allows the student to interact with the ITS. It most commonly involves a computer screen, keyboard, and mouse (Burton, 1982; Duchastel, 1989; Epstein & Hillegeist, 1990; Freedman, 2000).

The expert module contains all of the knowledge or information that is to be learned by the student, similar to an answer key. It is created with the assistance of an expert in the area that the ITS is designed to teach (Burton, 1982; Duchastel, 1989; Epstein & Hillegeist, 1990; Freedman, 2000). For example, the most reliable empirical research was used in the creation of the expert module of the ITS for deception detection and investigative interviewing used in the current studies.

The student module contains descriptions of the student's behaviors and knowledge. By comparing the student module to the expert module, gaps in the student's knowledge can be identified (Burton, 1982; Duchastel, 1989; Epstein & Hillegeist, 1990; Freedman, 2000). For example, in the current studies a student may be basing his or her deception judgments on non-verbal cues or asking direct (yes/no) questions to interviewees. However, the expert module states that verbal cues to deception are preferable to non-verbal cues and open-ended questions are preferable to direct questions. When the student's incorrect behaviors are compared to the expert module's answer key, the student's misconceptions are identified.

The tutor module is then designed to address the student's misconceptions which were identified using the student and expert modules. It takes corrective action by providing feedback to the student regarding his/her performance and re-teaching misunderstood information (Burton, 1982; Duchastel, 1989; Epstein & Hillegeist, 1990; Freedman, 2000). For example, in the current study, the tutor module may require the student to review the difference between verbal and non-verbal cues to deception and open- and close-ended questions.

McArthur, Stasz, and Zmuidzinas (1990) describe how these four modules work together to train the student. Initially, the student conveys his/her knowledge of the area using the interface module. Next, the ITS assesses the student's current state of knowledge regarding a topic by comparing student responses to the expert module. The tutor module then sets specific goals for the current tutoring session. The ITS then organizes the current session and delivers it to the student using the interface module. Finally, the process is repeated as the student again conveys his/her knowledge of the area using the interface module. Each of these steps is repeated many, many times throughout the tutoring process as constant evaluation, learning, planning, and re-evaluation takes place.

### **Virtual Human Tutoring**

One of the more recent developments in intelligent tutoring is the use of virtual humans. A virtual human is the creation of a human being in image and voice using computer-generated imagery and sound (Magnenat-Thalmann & Thalmann, 2004). Using a virtual human has many of the advantages of actual human tutoring while eliminating several disadvantages (Brooks, 2002). For example, human beings learn better when interacting face-to-face with another human being (e.g., having a conversation in person) than when no interaction takes place (e.g., reading a book). By using a virtual human, this one-on-one interaction can still take place, but

without requiring that an actual human tutor be present. Designing and replicating one virtual human is also much easier and cheaper than employing many human beings.

Using a virtual human to train deception detection will also avoid several pitfalls that have been problematic for other deception detection training. For example, other training approaches often involve human “role playing.” During this role playing, human beings find it difficult to consistently exude appropriate cues to deception, provide proper feedback, and/or behave in the same standardized, controlled way for each participant. A virtual human has no such problem.

The ITS for the current studies involves an experiential component (a virtual human, as opposed to a passive learning environment), a coaching component (that guides the learner through the interview experience), a reflective tutor component (that assesses learning, and provides reflective feedback on behaviors), and ultimately feedback on performance (at the end, learners will be provided a summary of the vHuman’s behaviors indicating lying/truthfulness). By using a virtual human intelligent tutoring system to train deception detection and investigative interviewing techniques, the current study is cheaper, faster, and more standardized than one-on-one human tutoring. The system also retains the advantages of student interaction with a human tutor that are sacrificed when using more traditional computer-based training approaches.

### **Overview of the Proposed Study**

Frank and Feeley (2003) identified six criteria necessary in training studies for deception detection. Although it is difficult for a single study to meet all six criteria, the current study employs a number of these criteria to improve upon past deception detection training studies. For example, two of the criteria involve ensuring that deception detection training programs are

relevant to real-world settings and that they generalize across deception detection situations. Compared to a university laboratory the virtual human intelligent tutoring program used in the current study can more easily imitate real-world settings, as well as provide multiple scenarios that seek to generalize across scenarios/settings. Another criterion is that participants must receive proper training. In the present research, training involved those cues to deception that have an empirical basis for diagnosticity, as well as effective interviewing techniques to elicit these cues. A final criterion is that deception detection training studies conduct proper pre-testing and post-testing of deception detection performance. The current research used standardized stimuli and data collection procedures, as well as pilot studies, to ensure this criterion was met.

The goal for the current research was to assess the effectiveness of two training programs at improving participants' ability to elicit and detect accurate cues to deception. Participants in the vHITS training condition were trained using a virtual Human Intelligent Tutoring System. This program trained participants to use investigative interviewing techniques to elicit reliable verbal cues to deception from an interviewee and to then accurately detect deception using these cues. Participants in the Computer Based Training (CBT) condition completed a comparable training program that did not involve any virtual human intelligent tutoring system. Participants in a control condition received no training. Training conditions were then compared to the no-training condition on a deception detection task and an investigative interviewing task.

Two key research questions were addressed. First, can appropriate training (on diagnostic cues/approaches) improve deception detection performance? It was hypothesized that participants in both training conditions would perform better on the deception detection and investigative interviewing post-tests than those in the no-training control condition. Second, can

interactive vHuman technology outperform a more passive training environment? It was hypothesized that participants in the vHITS condition would perform better on both post-tests than those in the CBT condition.

## METHOD

### Participants

One hundred and five undergraduate students from the University of Texas at El Paso (UTEP) were recruited to participate in this experiment in exchange for credit towards a research requirement of introductory psychology classes at UTEP. All 105 participants completed the deception detection task at pre-test and post-test. However, 30 of the 105 participants also completed the investigative interviewing tasks at pre-test and post-test. In addition to the 105 participants, 60 additional undergraduate students from UTEP were recruited to act as interviewees during the pre-test and post-test investigative interviewing task in exchange for research credit. More detailed descriptions of both the deception detection task and the investigative interviewing task are provided below.

### Design

A 3 x 2 mixed factorial design was employed in which participants were randomly assigned to one of three between-subject conditions: a virtual Human Intelligent Tutoring System (vHITS) condition, a Computer Based Training (CBT) condition, or a no-training control condition. All three conditions completed both a pre-test and post-test assessment (repeated measure).

### Materials

**Pre-test and post-test stimuli.** The stimulus materials used to evaluate participants' deception detection accuracy consisted of previously collected true and false alibi statements. In these statements, individuals provided either a truthful or deceptive account of their whereabouts

three nights prior to the interview. Participants were randomly assigned to either the truthful or deceptive condition.

Participants in the truthful condition were given the following instructions:

“Three days ago a crime took place between the hours of 7:00 PM and 10:00 PM. The police have targeted you as a suspect for this crime. You will be asked to provide a detailed true alibi of where you were and what you were doing at the time of the crime. You will have ten minutes to remember as many details as you can about your whereabouts.”

Participants in the deceptive condition were given slightly different instructions:

“Three days ago a crime took place between the hours of 7:00 PM and 10:00 PM. The police have targeted you as a suspect for this crime. You will be asked to provide a detailed false alibi of where you were and what you were doing at the time of the crime. We will give you a scenario for what you were doing and you will have ten minutes to create or fabricate as many details as you can about your whereabouts. Remember this is a false alibi, so don’t simply relate what you actually did on another night. Make your story a false experience. For example, DO NOT just tell us what you did Thursday night instead of Friday night. Make this a novel false experience.”

Participants in the deceptive condition were then assigned to one of following 10 scenarios by the experimenter: (1) movie theatre, (2) dinner and drinks, (3) bowling, (4) rent a movie, (5) card game, (6) concert, (7) athletic event, (8) casino, (9) video games, or (10) barbeque/ house party. Participants in both conditions were then given 10 minutes to prepare their alibi statements, after which they were videotaped providing their alibis.

The video clips of the alibi statements begin with the interviewer asking the suspect, “Where were you on the night in question between the hours of 7:00 PM and 10:00 PM?” The

participants would then provide a narrative for their whereabouts after which the interviewer would ask several follow-up questions based on the details of the participant's statement.

Interviews ranged from approximately 1.5 to 7.0 min ( $M = 2:47$ ). Forty-two videos (21 true and 21 false) were collected in total. Eighteen participants then pilot tested the videos by viewing them and attempting to distinguish between true alibis and false alibis. The order in which the videos were presented was randomized across participants. Accuracy rates for each of the forty two videotaped alibi statements were averaged across all 18 participants. Ten true and ten false alibis were selected for inclusion in the present experiment by eliminating those statements associated with the most extreme accuracy rates. Accuracy for the 20 videotaped alibi statements selected for inclusion averaged 53.59%. This overall accuracy rate is very similar to the overall accuracy rate reported in the DePaulo et al. (2006) meta-analysis of 54%.

From these 20 videotaped alibi statements, two comparable sets of 10 alibi statements (5 true, 5 false) were created for pre-testing and post-testing purposes in the study. The selection of alibi statements to each set was based on the discrimination accuracy for each individual alibi statement mentioned above so that the final two sets of 10 alibi statements had equivalent discriminability. During pre-testing and post-testing, the presentation of these videos was randomized within each set to prevent order effects.

Participants were also asked to complete a 16-item questionnaire designed to test their knowledge of cues to deception and interviewing tactics (see Appendix A). Questions 1-6 tested students on the 6 verbal cues taught in training. Questions 7-12 tested students on the 6 non-verbal cues taught in training. Questions 13-14 tested students on the 2 investigative interviewing techniques to be taught in training. Finally, questions 15-16 tested students on the 2

accusatory interviewing techniques to be taught in training. Each of these cues and techniques will be further explained below.

Participants completed the pre-testing and post-testing via a PC computer that was used to display the video clips on a 19" LCD monitor. Each participant wore headphones to receive the audio portion of the videos. MediaLab software was used to display the videos to participants and to collect responses following each video statement.

**vHITS training program.** The virtual Human Intelligent Tutoring System (vHITS) was created in cooperation with the Institute of Creative Technologies at the University of Southern California (see <http://ict.usc.edu>). This program was designed to train students to elicit and identify accurate cues to deception through interactions with a virtual human. The program can be broken down into three components: (1) the tactical questioning program, (2) the avatar, and (3) the intelligent tutor. The tactical questioning program allows the student to ask almost any question he/she wants and receive a plausible response that may contain diagnostic cues to deception. The avatar allows the student to interact with a virtual human. Finally, the intelligent tutor teaches students to ask rapport-building and open-ended questions, to pay attention to diagnostic verbal cues to deception, and to disregard non-diagnostic non-verbal cues to deception. These three components of vHITS allow a learning process to take place in which constant evaluation, teaching, learning, and re-evaluation takes place.

The *tactical training program* (TACQ) enabled the authors of vHITS to create a tutoring program that would allow participants to interview virtual humans and receive appropriate responses within the context of two specific crime scenarios. The first scenario is the bombing of an abortion clinic and the second scenario involves a murder investigation. The virtual human character interviewed in the first scenario is a man who was picked up on the side of the road

near the crime scene. The virtual human character interviewed in the second scenario is a woman who was an employee of a pawn store owner who had been murdered. In both scenarios, the virtual human character claims to have witnessed but not participated in the crimes. The task of the student is to interview the virtual human character to determine if he/she is lying or telling the truth. If the suspect is lying, then he/she is a criminal that was involved in the crime, but if the suspect is telling the truth, then he/she is just a witness that can provide helpful information. Participants completed both interview scenarios and the presentation of the characters/scenarios was counterbalanced.

The program is set up such that participants type questions into a text box and the virtual human speaks the response. After the participant types his/her question into the text box, a list of the top 5 closest matching questions is provided to the participant. In order for the TACQ program to successfully recognize all different forms of a question, many variations of the same question were authored into the system. This feature ensures that even in a situation where a question does not have an exact match, the participant will be able to ask a similar question. Questions about almost anything an interviewer might want to know were programmed into vHITS including questions regarding: details about the character's personal life (e.g. family, place of residence, criminal history, etc.), beliefs (political opinions, stance on abortion, etc.), the incident (what they claim to have seen, other witnesses, etc.), his/her alibi, and persons who might be responsible for the incident.

Deceitful and truthful responses to each question were also created. Based on deception detection research, six diagnostic verbal cues to deception were identified and included in the virtual human's responses. A description of how each of these six verbal cues is manifested in deceitful vs. true responses follows: (1) logical structure – deceitful responses contain less

logical structure than truthful responses; (2) spontaneous corrections – deceitful responses contain fewer spontaneous corrections than truthful responses; (3) admitted lack of memory – deceitful responses contain fewer instances of admitted lack of memory than truthful responses; (4) detail – deceitful responses contain fewer details than truthful responses; (5) temporal and spatial detail – deceitful responses contain fewer temporal and spatial details than truthful responses; and (6) negative statement/complaints – deceitful responses contain more negative statements/complaints than truthful responses. Each of these six verbal cues to deception were found to have large effect sizes for significant differences between liars and truth tellers in the DePaulo et al. (2003) meta-analysis. Several of these criteria are also included in the Criterion-Based Content Analysis (CBCA) technique and several Reality Monitoring (RM) studies.

Including each of the six verbal cues in each virtual human response would make the responses seem unrealistic when compared to actual human responses in the real world. Therefore, each virtual human response does not include all of the cues available. Instead, responses were tailored to what the authors determined was a plausible deceitful or truthful response to each question. For questions unrelated to the crime (e.g., “How old are you?” or “What can you tell me about your family?”), the deceitful and truthful responses differ very little or not at all. However, for crime-related questions (e.g., “What did you witness this morning?”), deceitful and truthful responses differ a great deal. These diagnostic, verbal cues to deception are also linked to the interviewing style of the student. If the student asked the suspect an open-ended or rapport-building question, he/she elicited more of the accurate verbal cues to deception from the suspect. On the other hand, if the student asks a closed-ended or accusatory question, he/she elicited fewer accurate verbal cues to deception from the suspect.

Based on the deception detection literature, six non-verbal cues to deception were also identified and included in the virtual human's responses. These six non-verbal cues are as follows: (1) gaze aversion (character looks away), (2) closed posture (character crosses arms), (3) tics (characters rubs head/neck/chin), (4) gestures (character signals me/you/left/right), (5) hand and arm movements (character makes a chopping motion with arms), and (6) head movements (character shakes/nods head). Each of these non-verbal cues to deception is either non-diagnostic or in the opposite direction of commonly held beliefs regarding deception detection (DePaulo et al., 2003; Vrij, 2008). Therefore they were included equally in both the truthful and deceptive responses to serve as distracters to participants. Again, each of the six non-verbal cues is not present in each virtual human response, but rather responses were tailored to what the author determined was a plausible response to each question.

The *avatars* used in vHITS were provided by one of ICT's current projects entitled Virtual Patient. A young adult female avatar and a young adult male avatar were chosen to match the characters in the scenarios. These avatars are linked to the TACQ program to provide the student an interactive experience with a virtual human providing both diagnostic and non-diagnostic cues to deception.

The third component of vHITS is the *intelligent tutor*. Specific points in the interview where tutoring would be most beneficial were identified. At these points interaction with the virtual human is paused and the tutor appears on the screen, providing feedback on the student's current performance and advice on investigative interviewing or deception detection. For investigative interviewing, this feedback and advice focused on rapport building, open-ended questioning, and staying away from accusations. For example, the student may begin the interview by immediately asking the suspect questions regarding the incident. In this case the

tutor would give the following hint regarding building rapport to the student: “Before jumping into issues regarding the incident, it is a good idea to get to know the suspect by asking questions about him/her and his/her background and beliefs. This allows you to learn about the individual and his/her motivations, as well as makes the suspect feel more comfortable in the interview setting.”

For deception detection, this feedback and advice focuses on encouraging the student to attend to the diagnostic verbal cues to deception and to ignore the non-diagnostic non-verbal cues. For example, the student may ask the suspect a question to which the truthful response contains more detail. In this case the tutor would give the following hint to the student: “Pay attention to the amount of details in the suspect’s answer. True statements tend to be more detailed than false statements.” On the other hand, the student may ask the suspect a question to which the deceitful response contains more negative statements and complaints. In this case the tutor would give the following hint to the student: “Pay attention to the number of negative statements and complaints in the suspect’s answer. False statements tend to contain more negative statements and complaints than true statements.”

**CBT training program.** A Computer-Based Training (CBT) program was also created to provide a comparable training condition that mimics a more standard learning environment. Consistent with the vHITS program, the CBT program was designed to train students to elicit and identify accurate cues to deception. Both programs trained the same investigative interviewing techniques and cues to deception to participants. The primary difference between vHITS and CBT is that the latter involved a more passive learning environment while the former provided for one-on-one interaction with a virtual human intelligent tutor. The CBT program was broken down into three components: (1) a written presentation, (2) videotaped examples,

and (3) a vocal explanation of the written presentation. These three components were combined into one Windows Media Video (WMV) file that was played on a computer and observed by the student. The CBT program was very similar to a typical classroom lecture with the exception that the student was not be able to interact with the instructor. The student saw the Microsoft Office PowerPoint presentation on the computer monitor and heard a vocal explanation of the concepts on the screen through headphones. After each concept was thoroughly identified and explained, the student saw a videotaped example within an interview setting.

The *written presentation* was created using Microsoft PowerPoint and provided the learner with key concepts and definitions. It contained instruction regarding both cues to deception and interviewing techniques. The same six diagnostic verbal cues that were included in the vHITS program were also included in the CBT program, including logical structure, spontaneous corrections, admitted lack of memory, level of detail, temporal and spatial cues, and negative statements/complaints. In addition, the same six non-diagnostic non-verbal cues that were included in the vHITS program were also included in the CBT program, including: gaze aversion, closed posture, tics, gestures, hand and arm movements, and head movements. Finally, the interviewing techniques taught were also the same as those in the vHITS program, such that students were instructed to ask questions that were open-ended and/or built rapport with the interviewee and to avoid questions that were close-ended and/or accusatory.

The second component of the CBT program is the *videotaped examples of the concepts* taught in the written presentation. These examples were created to demonstrate the particular cue to deception or interviewing approach that was being taught. For example, mock interviews were videotaped in which an interviewer and interviewee acted out a scene taken from the scenarios developed for vHITS that demonstrated a particular cue to deception or provided

examples of open-ended vs. close-ended interviewing strategies. By controlling the content of the demonstration videos across the CBT and vHITS condition, it was hoped that participants might be provided a comparable training experience.

The final component of the CBT program was the *vocal explanation of the written presentation*. For this component of the training, the researcher lectured using the written presentation created in Microsoft Office PowerPoint as a guide or outline. This vocal lecture included in-depth explanations regarding each cue to deception identified in the written presentation. The vocal presentation also helped ensure that learners attended to the material presented.

Taken together, the CBT training was very similar to the vHITS training with a few important differences. Both programs trained the same interviewing techniques and cues to deception and implemented the same two scenarios in the process of training. However, the CBT program provided a more passive training environment in which the student did not directly interact with the tutor in any way. On the other hand, the vHITS program provided a more interactive training environment in which the student was tutored one-on-one and provided an opportunity to implement learned skills by interacting with a virtual human who could respond in a more naturalistic environment.

**No-training control filler task.** A filler task was provided to the no-training control condition (during the time that the two training conditions completed their training on deception detection and interviewing). Similar to the two training programs, this task took participants just under an hour to complete. Specifically, participants completed a recognition memory experiment for human faces in which they were presented with a sequence of faces to study,

followed by a recognition test. As such, the filler task was in no way related to deception detection or investigative interviewing.

## **Procedure**

The current experiment consisted of three phases that were completed over a five day period. In the following paragraphs, the participant's actions during each of these three phases is described in detail.

**Phase 1: Pre-testing.** When a participant arrived on Day 1, he/she completed Phase 1 of the experiment. Upon arriving at the lab, the participant was presented with a consent form which they read and signed prior to participation in the study. The participant was then randomly assigned to one of the following three conditions within the study: vHITS training, CBT training, or no-training control.

After being assigned to a condition, 30 of the 105 participants were randomly selected to participate in an investigative interviewing task. The investigative interviewing task that participants completed was very similar to the method used to collect the stimulus materials used to evaluate participants' deception detection accuracy. First, participants were recruited and asked as interviewees to provide either a truthful or deceptive account of their whereabouts three nights prior to the interview. These interviewees were randomly assigned to either a truthful or deceptive condition.

Participants in the truthful condition were given the following instructions:  
“Three days ago a crime took place between the hours of 7:00 PM and 10:00 PM. The police have targeted you as a suspect for this crime. You will be asked to provide a detailed true alibi of where you were and what you were doing at the time of the crime. You will have ten minutes to remember as many details as you can about your whereabouts.”

Participants in the deceptive condition were given slightly different instructions: “Three days ago a crime took place between the hours of 7:00 PM and 10:00 PM. The police have targeted you as a suspect for this crime. You will be asked to provide a detailed false alibi of where you were and what you were doing at the time of the crime. We will give you a scenario for what you were doing and you will have ten minutes to create or fabricate as many details as you can about your whereabouts. Remember this is a false alibi, so don’t simply relate what you actually did on another night. Make your story a false experience. For example, DO NOT just tell us what you did Thursday night instead of Friday night. Make this a novel false experience.”

Participants in the deceptive condition were then randomly assigned to one of following 10 scenarios by the experimenter: (1) movie theatre, (2) dinner and drinks, (3) bowling, (4) rent a movie, (5) card game, (6) concert, (7) athletic event, (8) casino, (9) video games, or (10) barbeque/ house party. Participants in both conditions were then given 10 minutes to prepare their alibi statements, after which they were videotaped as they provided their alibis to the participants acting as interviewers.

The 30 participants acting as interviewers were given the following instructions: “Three days ago an imaginary crime took place between the hours of 7:00 PM and 10:00 PM. The police have apprehended a suspect for this crime. Your task is to act as a police investigator and conduct a mock interview of this suspect regarding his/her whereabouts on the date and time in question. At the end of the interview you will also decide whether the suspect was lying or telling the truth. You may ask the suspect any questions you desire during your mock interview.”

The interviewer was then videotaped as he/she questioned the interviewee. These interviews lasted between one and ten minutes and were later coded for various interviewing techniques. After completing the interview, the interviewee was thanked for their assistance and dismissed, while the interviewer provided a veracity judgment and estimated his/her confidence in the judgment via a 0 to 100% scale (0% = *not confident at all*, 100% = *extremely confident*).<sup>1</sup>

Following the completion of this investigative interviewing task by a randomly selected group of 30 participants, all 105 participants were seated in front of a computer monitor as the experimenter began the deception detection task portion of the pre-test. Participants were asked to view a series of videotaped interviews involving truthful and deceitful alibi statements. The following instructions were provided: “Several people were videotaped giving alibi statements for their whereabouts on a specific date and time. Some of these people are giving true alibis and some are giving false alibis. Your job is to watch the videos of these alibis and identify which people are lying and which people are telling the truth. The video of the person's alibi will play on the computer. At the end of each video a new screen will appear asking you to indicate if the suspect is lying or telling the truth and to provide an estimate of your confidence in this decision.”

All participants then viewed the series of 10 interviews. Following each interview, participants were asked to provide a veracity judgment and to estimate their confidence via a 0 to 100% scale (0% = *not confident at all*, 100% = *extremely confident*). The presentation of videos was randomized and counterbalanced across participants to control for any potential order effects.

Following the presentation of all 10 video clips, participants were asked to complete a 16-item questionnaire assessing their knowledge of interviewing techniques and cues to

deception. After completing the questionnaire, participants were thanked for their assistance and informed of when they needed to return to continue the experiment.

**Phase 2: Training on cues to deception and interviewing approaches.** Participants returned two days after the pre-test, on Day 3, to complete Phase 2 of the experiment. During this phase of the experiment, participants in the vHITS and CBT training conditions completed the appropriate training program described above, while the control condition completed an unrelated filler task involving the perception and recognition of faces. Each training program lasted approximately one hour. After completing the training, participants were again thanked for their assistance and informed of when they needed to return to complete the experiment.

**Phase 3: Post-testing.** Finally, all participants returned four days after the pre-test, on Day 5, to complete the third and final phase of the experiment. During this phase of the experiment, participants completed the post-test. The post-test was identical to the pre-test. Specifically, the 30 participants that were randomly selected to complete the investigative interviewing task again completed that task with a different interviewee who had been randomly assigned to lie or tell the truth. In addition all 105 participants completed an assessment of their deception detection performance via the presentation of 10 truthful or deceitful alibi statements (note that these videos were not viewed previously) as well as the 16-item questionnaire assessing their beliefs regarding cues to deception and interviewing approaches. After completing the post-test, participants were debriefed, thanked for their assistance, and dismissed.

## RESULTS

### Deception Detection Performance

Participants were asked to provide judgments of truth or deception and to estimate their confidence in this decision via a 0% to 100% scale. Signal detection estimates were used to analyze the data in order to quantify participant's ability to discern between signal (e.g., diagnostic cues to deception) and noise (e.g., non-diagnostic behaviors unrelated to deception) in a statement (Lu & Eskew, 2007). Signal detection theory (SDT) is used to measure the way decisions are made under conditions of uncertainty (Mueller & Weidemann, 2008), such as detecting deception within an alibi statement (Meissner & Kassin, 2002). SDT assumes that the decision maker is not a passive receiver of information, but an active decision-maker who makes difficult perceptual judgments (Starns et al., 2008).

Participant responses were separated into one of the following four categories: (1) hits (the proportion of true statements correctly identified as such), (2) misses (the proportion of true statements incorrectly identified as false), (3) correct rejections (the proportion of false statements correctly identified as such), and (4) false alarms (the proportion of false statements incorrectly identified as true; Mueller & Weidemann, 2008). These hit, miss, correct rejection, and false alarm rates were then used to compute signal detection estimates of discrimination accuracy ( $A_z$ ) (cf. Snodgrass & Corwin, 1988; Verde, MacMillan, Rotello, 2006).  $A_z$  was computed as:

$$(1) A_z = \Phi \left( \frac{d'}{\sqrt{2}} \right),$$

where  $d' = z_H - z_{FA}$ ,

and response bias (C) was computed as:

$$(2) C = 0.5 (z_{FA} + z_H).$$

Theoretically, discrimination accuracy and bias are independent constructs (Lu & Eskew, 2007; Snodgrass & Corwin, 1988). Analysis of the observed correlations between discrimination accuracy and response bias at pre-test ( $r = -.11$ , *ns.*) and post-test ( $r = -.15$ , *ns.*) confirmed their independence. Discrimination accuracy refers to participant's ability to detect whether a signal ("deception") is present or absent amidst the noise of background events (Starns et al., 2008). Response bias represents the extent to which one response ("truth" or "deception") is more probable than another (Starns et al., 2008).

Table 1 provides mean and standard deviation estimates for performance on the deception detection task (hits, false alarms,  $A_z$ , C, and confidence) across training conditions and pre-test/post-test. A 3 (between-subjects measures: vHITS vs. CBT vs. control) x 2 (within-subjects measures: pre-test vs. post-test) mixed Analysis of Variance (ANOVA) assessed the influence of training on *discrimination accuracy* ( $A_z$ ). A main effect of repeated testing was observed,  $F(1,102) = 18.73$ ,  $p < .001$ ,  $\eta_p^2 = 0.16$ , such that participants improved their performance from pre-test to post-test. A significant interaction was also observed,  $F(2,102) = 3.31$ ,  $p < .05$ ,  $\eta_p^2 = 0.06$ . Follow-up pairwise comparisons indicated that both the Computer Based Training (CBT),  $t(34) = 3.88$ ,  $p < .001$ ,  $d = 0.62$ , and the virtual Human Intelligent Tutoring System (vHITS),  $t(34) = 3.37$ ,  $p < .01$ ,  $d = 0.46$ , training conditions significantly improved deception detection performance. In contrast, the control condition showed no significant learning effect across the pre- and post-test,  $t(34) = 0.39$ , *ns.*,  $d = 0.06$ . Figure 1 provides a depiction of the mean *discrimination accuracy* ( $A_z$ ) scores at both pre-test and post-test for all conditions.

A 3 (between-subjects measures: vHITS vs. CBT vs. control) x 2 (within-subjects measures: pre-test vs. post-test) mixed ANOVA was also used to assess the influence of training on *response bias* (C). A main effect of repeated testing was observed,  $F(1,102) = 47.39$ ,  $p <$

.001,  $\eta_p^2 = 0.32$ , indicating that participants' response bias became conservative at post-test (i.e., more likely to respond "truth"). A marginally significant interaction was observed,  $F(2,102) = 2.70, p < .08, \eta_p^2 = 0.05$ . Follow-up pairwise comparisons indicated that both the CBT,  $t(34) = 4.85, p < .001, d = 0.76$ , and the vHITS,  $t(34) = 4.04, p < .001, d = .62$ , training conditions significantly increased participants' likelihood to judge a statement as truthful. Participants in the control condition,  $t(34) = 2.82, p < .01, d = 0.45$ , also showed a significant increase in likelihood to judge a statement as truthful. Figure 2 provides a depiction of the mean *response bias (c)* scores at both pre-test and post-test for all conditions.

Participants estimates of confidence were similarly analyzed using a 3 (between-subjects measures: vHITS vs. CBT vs. control) x 2 (within-subjects measures: pre-test vs. post-test) mixed ANOVA, resulting in a main effect of repeated testing,  $F(1,102) = 22.16, p < .001, \eta_p^2 = 0.18$ , such that participants expressed greater confidence in their decisions at post-test. A significant interaction was also observed,  $F(2,102) = 3.37, p < .05, \eta_p^2 = 0.06$ . Pairwise comparisons demonstrated that both the CBT,  $t(34) = 2.10, p < .05, d = 0.19$ , and the vHITS,  $t(34) = 4.98, p < .001, d = 0.47$ , conditions significantly increased their confidence as a function of training, while the control condition showed no significant effect,  $t(34) = 1.19, ns., d = 0.09$ . Figure 3 provides a depiction of the mean confidence scores at both pre-test and post-test for all conditions.

Participants' learning of the validity of various cues to deception and interviewing techniques as a function of training condition was also assessed. Table 2 provides mean and standard deviation estimates for performance on the learning questionnaire (verbal cues, nonverbal cues, and interviewing skills) across training conditions and pre-test/post-test. First, a 3 (between-subjects measures: vHITS vs. CBT vs. control) x 2 (within-subjects measures: pre-

test vs. post-test) mixed ANOVA on participants' knowledge of diagnostic verbal cues to deception demonstrated a significant main effect of repeated testing,  $F(1,102) = 143.79, p < .001, \eta_p^2 = 0.59$ , such that participants demonstrated improved knowledge at post-test. A significant main effect of training condition was also observed,  $F(2,102) = 12.72, p < .001, \eta_p^2 = 0.20$ , such that both training conditions demonstrated greater performance on the task than did the control condition ( $ps < .001$ ). In addition, a significant interaction was observed,  $F(2,102) = 22.73, p < .001, \eta_p^2 = .31$ . Pairwise comparisons showed that both the CBT,  $t(34) = 9.16, p < .001, d = 1.45$ , and the vHITS,  $t(34) = 9.07, p < .001, d = 1.49$ , conditions significantly improved their knowledge of diagnostic verbal cues to deception, while the control condition showed no significant learning effect,  $t(34) = 1.65, ns., d = 0.22$ . Figure 4 provides a depiction of the mean scores for knowledge of diagnostic verbal cues to deception at both pre-test and post-test for all conditions.

Next, a 3 (between-subjects measures: vHITS vs. CBT vs. control) x 2 (within-subjects measures: pre-test vs. post-test) mixed ANOVA was conducted on participants' knowledge of non-diagnostic nonverbal cues to deception. Results demonstrated a significant main effect of repeated testing,  $F(1,102) = 50.62, p < .001, \eta_p^2 = 0.33$ , such that participants demonstrated improved knowledge at post-test. A significant main effect of training condition was also observed,  $F(2,102) = 7.81, p = .001, \eta_p^2 = 0.13$ , such that both training conditions demonstrated greater performance on the task than did the control condition ( $ps < .01$ ). A significant interaction was also observed,  $F(2,102) = 7.38, p < .01, \eta_p^2 = .13$ . Pairwise comparisons showed that both the CBT,  $t(34) = 4.49, p < .001, d = 0.61$ , and the vHITS,  $t(34) = 5.25, p < .001, d = 0.81$ , conditions significantly improved their knowledge of non-diagnostic non-verbal behaviors, while the control condition showed only a marginally significant learning effect,  $t(34)$

= 1.96,  $p < .06$ ,  $d = 0.22$ . Figure 5 provides a depiction of the mean scores for knowledge of non-diagnostic nonverbal behaviors at both pre-test and post-test for all conditions.

Finally, a 3 (between-subjects measures: vHITS vs. CBT vs. control) x 2 (within-subjects measures: pre-test vs. post-test) mixed ANOVA on participants' knowledge of investigative interviewing skills demonstrated a significant main effect,  $F(1,102) = 100.63$ ,  $p < .001$ ,  $\eta_p^2 = 0.50$ , such that participants demonstrated improved knowledge at post-test. A significant main effect of training condition was also observed,  $F(2,102) = 28.03$ ,  $p < .001$ ,  $\eta_p^2 = 0.36$ , such that both training conditions demonstrated greater performance on the task than did the control condition ( $ps < .01$ ). A significant interaction was also observed,  $F(2,102) = 16.46$ ,  $p < .001$ ,  $\eta_p^2 = 0.24$ . Pairwise comparisons showed that both the CBT,  $t(34) = 6.47$ ,  $p < .001$ ,  $d = 0.97$ , and the vHITS,  $t(34) = 9.10$ ,  $p < .001$ ,  $d = 1.37$ , conditions significantly improved their knowledge of investigative interviewing techniques, while the control condition showed no significant learning effect,  $t(34) = 1.35$ , *ns.*,  $d = 0.17$ . Figure 6 provides a depiction of the mean scores for knowledge of investigative interviewing techniques at both pre-test and post-test for all conditions.

### **Interviewing Performance**

Participants' videotaped interviews were coded by three research assistants for the frequency of the following four interviewing behaviors: (1) rapport-building questions, (2) open-ended questions, (3) accusations, and (4) close-ended questions. These frequencies were then averaged and, similar to the above analyses, a series of 3 (between-subjects measures: vHITS vs. CBT vs. control) x 2 (within-subjects measures: pre-test vs. post-test) mixed ANOVA's assessed differences between pre-test and post-test for the four interviewing behaviors. Only one significant effect was observed. A power analysis suggested that despite the small sample size

( $N = 10$  per condition), there were no significant differences in effect size between the two training conditions,  $\eta_p^2 = 0.01$ , suggesting that a failure to find differences was not due to insufficient power. Of the four interviewing behaviors, significant differences between pre-test and post-test were observed only for accusations, suggesting that trained participants were less likely to make accusations during the post-test than during pre-test. However, this effect may be considered spurious given the extremely low frequency of accusations at both pre-test ( $M = 1.63$  accusations across all conditions) and post-test ( $M = 0.60$  accusations across all conditions), as well as the low inter-rater reliability for this variable at both pre-test (average  $r = .24$ ) and post-test (average  $r = -.01$ ). Inter-rater reliability was good to excellent for all other pre-test variables (average  $r$  for rapport-building questions =  $.66$ , open-ended questions =  $.85$ , and close-ended questions =  $.67$ ) and post-test variables (average  $r$  for rapport-building questions =  $.76$ , open-ended questions =  $.74$ , and close-ended questions =  $.57$ ).

Due to the exploratory nature of this portion of the study, data were examined descriptively for trends, while assessing observed effect sizes. Two variables may be worth considering: frequency of rapport-building questions and average question/response length. Average question/response length was calculated by taking the total time of the interview in minutes and dividing it by the total number of questions asked in the interview. Results showed a trend towards: (1) interviewers in the two training conditions (CBT and vHITS) asking more rapport-building questions at post-test than the control condition (Control:  $M = 0.23$ , CBT:  $M = 0.25$ , vHITS:  $M = 0.27$ ;  $\eta_p^2 = .04$ ) and (2) average question/response length at post-test being longer for the two training conditions (CBT and vHITS) than for the control condition (Control:  $M = 1.58$ , CBT:  $M = 4.30$ , vHITS:  $M = 3.88$ ;  $\eta_p^2 = .11$ ). Future research may be warranted to

assess these effects when a more robust training program is developed and a larger sample can complete this portion of the study.

## DISCUSSION

The goal of the current study was to assess two training programs designed to improve deception detection and investigative interviewing capabilities. Explanations of the results, weaknesses of the study, and future directions for this line of research will be discussed.

### **Deception Detection Performance**

Results of the current study suggested that both forms of training had an equally positive effect on participants' ability to detect deception when compared to a baseline (no training) control condition. This suggests that teaching participants to attend to diagnostic verbal cues to deception, as well as to ignore non-diagnostic nonverbal behaviors is an effective method for improving their deception detection performance. Participants' scores on the 16-item pre-test and post-test questionnaires confirmed that participants learned about cues to deception taught during training. When compared to the control condition, both training conditions showed large improvements from pre-test to post-test in the knowledge of both diagnostic verbal cues to deception and non-diagnostic non-verbal behaviors. As is common in other deception detection training studies, participants in both training conditions also showed an increase in confidence in their deception detection judgments after training.

It is important to note several additional aspects of the results. The increase in accuracy between the groups was only 6% for the CBT condition (pre-test accuracy = 51.0%, post-test accuracy = 57.0%) and about 8% for the vHITS condition (pre-test accuracy = 49.2%, post-test accuracy = 56.8%), resulting in an accuracy rate of approximately 57% for both training conditions. While this may seem small, past research has shown that 60% accuracy may represent a ceiling for improving deception detection performance. For example, Mann and Vrij (2006) and Vrij, Fisher, and Mann (2006) have shown that by increasing the cognitive load of the

liar, police officers were able to detect deception with approximately 60% accuracy. Porter, Woodworth, and Birt (2000) similarly demonstrated that after 5 weeks of training a group of Canadian federal parole officers were able to improve their deception detection accuracy rate about 14% to approximate the 60% accuracy mark.

These results also suggest there is no apparent value of interactive practice with feedback for the skill of detecting deception in recorded statements. It is common practice in deception studies to use recorded statements at pre-test and post-test. They are passive experiences that are easy to repeat and compare because participants simply watch the video and make a decision. In this study, the pre-test and post-test are similar in structure to the CBT condition (which consisted of a 40 min presentation on cues to deception with recorded examples). This may explain why no significant differences on accuracy were observed between the CBT and vHITS conditions at post-test. In many ways, the CBT condition had higher fidelity than the vHITS condition because the recorded statements used at both pre-test and post-test as well as in the CBT training show real people, using intonation, facial expressions, and so on. Given this, the results can even be considered encouraging since virtual human-based training was able to produce equivalent learning to a very strong passive training condition closely aligned to the test.

It is also of interest that the increase in accuracy for both training conditions was focused entirely on hits (judgments of truth) as opposed to correct rejections (judgments of deception). This increase in hits led to a response bias towards perceiving truth for both training conditions. One possible explanation for this truth bias may be the types of cues emphasized during training. There are two types of diagnostic cues to deception: (1) behaviors that are performed more often by truth tellers and (2) behaviors that are performed more often by liars (Vrij, 2008). For

example, truth tellers provide more details in their statements than liars, whereas liars make more complaints in their statements than truth tellers. (Bond & DePaulo, 2006).

The vast majority of the diagnostic verbal cues to deception that were emphasized in both training conditions involved behaviors performed more often by truth tellers than liars. Other deception techniques that emphasize these behaviors, such as Criteria-Based Content Analysis (CBCA) and Reality Monitoring (RM), also tend to result in a truth bias (Vrij, 2008). In contrast, the training that police receive, such as the Reid Technique, emphasizes behaviors that are supposedly performed more often by liars than truth tellers and results in a deception bias (Meissner & Kassin, 2002). Perhaps the reason for these biases is that participants trained in techniques such as CBCA, RM, and the training programs used in the current study ignore indicators that a statement is false because they are in search of indicators that the statement is true (e.g., verbal cues to deception), leading to a truth bias. In contrast police officers may ignore indicators that a statement is true (e.g., verbal cues to deception) because they are in search of indicators that the statement is false (e.g., non-verbal cues to deception), leading to a deception bias.

Using this logic, future research should focus on minimizing biases by focusing participants equally on diagnostic cues performed more often by truth tellers and those that are performed more often by liars. One possible way to accomplish this is to focus participants equally on diagnostic verbal and paraverbal cues to deception. A paraverbal cue to deception is any vocal behavior that accompanies speech, and research suggests that such cues are typically performed more often by liars than truth tellers. Some examples of diagnostic paraverbal cues include vocal uncertainty (e.g., ending statements with a higher tone as if asking a question), tension/nervousness (e.g., speaking with a higher pitch), rate of speech (e.g., speaking quickly),

and level of cognition (e.g., appearing to be thinking hard) (Sporer & Schwandt, 2006). Because biases may be correlated with the types of cues being examined, it is important that deception detection techniques rely on a variety of different types of diagnostic cues to deception rather than one specific type of cue.

It is important to distinguish between detecting deception and detecting cues to deception. The current study attempted to train participants to detect deception by focusing them on the most diagnostic verbal cues to deception (DePaulo et al., 2003). At pre-test and post-test, participants completed a deception detection task in which they distinguished between truth and false alibi statements. The videotaped alibi statements used at pre-test and post-test were pilot tested and chosen based on participants' discrimination accuracy, rather than being selected based on whether or not certain cues to deception were present in the videos. In contrast, both training programs involved teaching participants to detect and interpret diagnostic verbal cues to deception. The study attempted to determine whether knowledge of diagnostic cues to deception might transfer to a set of true and false statements that had not been screened for presence of the cues. Ideally, the act of detecting and interpreting accurate cues to deception should lead to better deception detection performance, as the results of this study indicated.

### **Interviewing Performance**

Results of the current study suggest that neither form of training had a positive effect on ability to conduct investigative interviews when compared to a baseline (no training) control condition. However, it is important to note several aspects of the results. The main focus of the study was to assess the deception detection component of the training, and therefore the investigative interviewing component was much more exploratory in nature. As a result, the sample size for this component of the study was small ( $N = 10$  per condition), resulting in low

power to detect effects. Still, a power analysis suggested that despite the small sample size ( $N = 10$  per condition), there were no significant differences in effect size between the two training conditions,  $\eta_p^2 = 0.01$ , suggesting that a failure to find differences was not due to insufficient power. When data was examined descriptively for trends, two trends of interest were observed: (1) interviewers in the two training conditions (CBT and vHITS) appeared to ask more rapport-building questions at post-test than the control condition and (2) the average length of response at post-test was longer for the two training condition (CBT and vHITS) than for the control condition. These trends are consistent with what is expected from a properly conducted investigative interview.

While results on the investigative interviewing task at pre-test and post-test showed no significant effect for training, it should be noted that participants' scores on the 16-item pre-test and post-test questionnaires confirmed that participants learned about proper investigative interviewing techniques taught during training. This result may suggest that while participants learned and understood the interviewing techniques being trained, they had difficulty applying these techniques during an actual interview. Perhaps the techniques trained were fundamentally different from the types of conversations one engages in on a daily basis. It may have been difficult to overcome habits towards a specific type of conversing developed over the course of years with just 40 min of focused training.

### **Format and Length of Training**

Also of interest is the length and format of the training. In the current study, undergraduate participants completed pre-test tasks on Day 1 for about an hour. Participants then returned to the lab two days later, on Day 3, and completed approximately one hour of training. Finally, participants returned to the lab two days later, on Day 5, and completed post-

test tasks for about an hour. This format and length for the training were chosen because of student schedules. Many undergraduate students have the same schedule on Monday, Wednesday, and Friday in which they attend classes with one or two hour breaks between classes. By conducting the study for one to two hour periods on Mondays, Wednesday, and Fridays, participating in the study was convenient for students. However, this format and length of training may not be ideal for two reasons.

First, the current format may have made retention of the training material difficult. If the training took place over the course of one or two days rather than five, retaining the training material may have been easier for participants. Second, the training session may not have been long enough for participants to obtain all the desired skills. If the training session lasted several hours, participants could interact more with the virtual human, ask more questions, and engage in more real-world scenarios.

Although the current training produced significant effects for deception detection performance and several learning measures, in many ways the format and length of the training made it difficult for participants to acquire the deception detection and investigative interviewing skills being taught. Future research should allow participants more training time and have less time between pre/post-test and training. Making these changes may result in better performance on the deception detection and investigative interviewing tasks as well as greater retention of the training material.

### **Issues with the vHITS Training**

The vHITS training program was designed to train participants to elicit and detect accurate cues to deception by interacting with a virtual human. The program was developed in just 4 months at the Institute for Creative Technologies at the University of Southern California.

Because the program was developed quickly, there were many issues with the program that might be improved and extended in future research.

The program was created using several existing programs at ICT. Unfortunately, these programs were not designed to work together resulting in several problems. First, vHITS was not a piece of software that could be installed easily on any computer. ICT was able to install the program on only one laptop that was then shipped to the University of Texas at El Paso (UTEP). Opening the program was another difficult undertaking, which took about 5 min for a trained graduate student. Several programs had to be opened in a specific order and some of the basic computer code had to be typed into the system for each participant. Once a participant finished interacting with the first virtual human, the program would have to be closed and restarted in order for the participant to interact with the second virtual human.

The display of vHITS could also be problematic. The left side of the computer screen contained a text box (where participants input their questions), a hint box (where participants could receive tips for conducting interviews and detecting deception), and a transcript of the interview taking place. The right side of the computer screen contained the virtual human avatar. If participants clicked anywhere on the right side of the computer screen the program would need to be closed and restarted. The programs used to run vHITS would run in the background as the participant interacted with the program. This was distracting to participants as over ten windows would be open at the bottom of the computer screen and clicking on any of them would cause the program to close.

The interactions between the participant and the virtual human were not always ideal. The program was set up so that when participants type questions into a text box, the five closest matches in the system to the participant's question are displayed. The participant then selected

the closest match to the question he/she wanted to ask and the virtual human spoke the response. Unfortunately, at times there was no close match to a participant's question. In these cases the participant would have to choose an unrelated question.

In addition, the voice of the virtual human was at times difficult to understand due to its awkward intonations and poor enunciation. One participant commented that it sounded "like the avatar was under water." Although the focus of the current study was on verbal and nonverbal cues to deception, future research may wish to focus on various paraverbal cues to deception. Several of these paraverbal cues were found to have large effect sizes for significant differences between liars and truth tellers in the DePaulo et al. (2003) meta-analysis, including: verbal/vocal uncertainty, tension/nervousness, and rate of speech. Unfortunately, the current avatar voices were not able to perform any of the diagnostic paraverbal cues to deception mentioned above. Future versions of vHITS should include the ability to perform these cues.

The amount of non-verbal behaviors that the virtual human was able to perform was also limited. In fact, none of the nonverbal behaviors that participants were taught during training (via hints appearing on the screen) could be performed by the avatar. Finally, only two virtual humans (one male and one female) and two scenarios (a bombing and a murder) were created. Additional characters and scenarios would be beneficial, and future research should focus on resolving these problems.

As virtual human technology matures, it will become easier to simulate human behavior with higher fidelity which will enable the system to address a greater range of novice misconceptions. For example, nervousness is commonly interpreted as a sign of deception when it is, in fact, not a reliable indicator (DePaulo et al., 2003). A nervous, but truthful, virtual human would provide an interactive example to demonstrate that there are often many causes of

nervousness, such as being asked questions. Beyond this, the two avatars fall into the category of question / answer virtual humans, meaning they do not possess a realistic model of the interaction with the user beyond the simple phase markers used by the tutor. Although sufficient for some learning goals, to tap more deeply into the investigative interview process, learning would likely benefit from models of emotions, proxemics, and consequences of what is being said (some of these aspects appear in other virtual human research; see Gratch et al., 2007; Swartout et al., 2006).

### **Potential for Application**

Despite its weaknesses, the vHITS system has the potential for future application with populations for whom deception detection is a useful skill (e.g., local, state, and federal law enforcement; Transportation Security Administration (TSA); and military). Plans are in place to test this training program at a United States Army installation in Fort Huachuca, Arizona. The program could also serve as part of the training for cadets (future police officers) at the El Paso Sheriff's Academy or as a periodic refresher for current law enforcement. Conducting training with these populations rather than undergraduate students may change the current results in two key ways.

With regard to deception detection performance, law enforcement officials are more likely to have a deception bias whereas lay individuals generally demonstrate a truth bias (Bond & DePaulo, 2006; Meissner & Kassin, 2002). Results of the present study show that the current training significantly increased participants' truth bias. If the effect of training on response bias remains consistent, training could alleviate law enforcement officials' deception bias rather than exacerbating bias as it did with the current participants' truth bias. This elimination of bias

would be beneficial for law enforcement as it may result in improved deception detection accuracy as well.

With regard to interviewing performance, many police officers are still trained primarily in accusatorial interrogation techniques (Vrij, 2008). Results of the present study show that the current training did not have a significant impact on interviewing behaviors. However, it is important to note that the baseline performance on the interviewing task of the current participants may be significantly better than the baseline performance of law enforcement officials. Results on the pre-test interviewing task showed that participants were already using more information-gathering interviewing techniques than accusatory interviewing techniques prior to receiving any training.

Averaged across all conditions and coders, 19% of participants' pre-test questions were rapport-building while just 8% were accusatory and 45% of participants' pre-test questions were open-ended while only 28% were closed ended. Because participants' pre-test performance was already high, obtaining significant effects may have been difficult. However, it is unlikely given the interviewing training that law enforcement officials currently receive that police officers' pre-test performance would be as high. Therefore, obtaining significant effects on the interviewing measures may be more likely with this population.

## **General Discussion**

Virtual human technology is in the early stages of being applied to the problem of teaching social interaction skills. The current examination of the vHITS training program could have been improved in several ways. Teaching participants not only about cues to deception performed more often by truth tellers than liars but also about those cues performed more often by liars than truth tellers would have been beneficial. The gap between what participants learned

and what they applied could have been bridged by allowing participants more training time and eliminating the delay between pre/post-test and training. Despite these limitations, results of the current study suggest that a virtual human-based system can increase learners' deception detection skills.

The question of whether programs like vHITS can outperform more passive training like CBT depends on whether the current program can be improved in several important ways. Future models of vHITS should focus on making the software easier to install and more user friendly. Interactions between the learner and virtual human could be improved by perfecting the question matching function or ideally eliminating it completely. Richer models of virtual human behavior would allow the program to better imitate humans. Finally, more scenarios and characters would allow learners to see how trained skills are applied to a wide variety of people and environments.

After making these improvements, vHITS can be applied to police and military environments. By teaching learners to ask rapport-building and open-ended questions rather than accusatory and closed-ended questions, vHITS will allow investigators to elicit accurate verbal cues to deception from interviewees. By focusing learners on diagnostic verbal cues to deception rather than nonverbal behaviors, vHITS will then allow investigators to detect the accurate cues to deception that they elicit from interviewees resulting in better deception detection performance.

## REFERENCES

- Akehurst, L., Bull, R., Vrij, A., & Köhnken, 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.
- Akehurst, L., Köhnken, G., Vrij, A., & Bull, R. (1996). Lay persons' and police officers' beliefs regarding deceptive behavior. *Applied Cognitive Psychology, 10*, 461-471.
- Anderson, J. R., Corbett, A. T., Koedinger, K. R., Pelletier, R. (1995) Cognitive tutors: Lessons learned. *Journal of the Learning Sciences, 4*(2), 167-207.
- Barrett, H. C., Frederick, D., Haselton, M., & Kurzban, R. (2006). Can manipulations of cognitive load be used to test evolutionary hypotheses? *Journal of Personality and Social Psychology, 91*, 513-518.
- Bond, C.F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality & Social Psychology Review, 10*(3), 214-234
- Brooks, L. (2002). Synthespians, Virtual Humans, and Hypermedia. In Veronica Hollinger and Joan Gordon, *Edging Into the Future: Science Fiction and Contemporary Cultural Transformation*. (pp. 57–59). University of Pennsylvania Press.
- Bull, R., & Milne, B. (2004). Attempts to improve the police interviewing of suspects. In: Lassiter, G. Daniel (Ed.), *Interrogations, confessions, and entrapment* (pp. 181-196). New York, NY, US: Kluwer Academic/Plenum Publishers.
- Burgoon, J. K., Blair, J. P., & Strom, R. E. (2008). Cognitive biases and nonverbal cue availability in detecting deception. *Human communication research, 34*, 572-599.
- Burton, R. R. (1982). Diagnosing bugs in a simple procedural skill. In: D. Sleeman & J. S. Brown (Eds.), *Intelligent Tutoring Systems* (pp. 157-183). New York: Academic Press.

- DePaulo, B. M., Lassiter, G. D., & Stone, J. I. (1982). Attentional determinants of success at detecting deception and truth. *Personality and Social Psychology Bulletin*, 8, 273-279.
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Charlton, K., Cooper, H., & Muhlenbruck, L. (2003). Cues to deception. *Psychological Bulletin*, 129(1), 74-118.
- DePaulo, B. M. & Morris, W. L. (2004). Discerning lies from truths: Behavioural cues to deception and the indirect pathway of intuition. In Granhag, Pär-Anders and Strömwall, Leif (Ed.), *The detection of deception in forensic contexts* (pp. 15-40). New York, Cambridge University Press.
- DePaulo, B. M., Rosenthal, R., and Rosenkrantz, J. (1982). Actual and perceived cues to deception: A closer look at speech. *Basic and Applied Social Psychology*, 3(4), 291-312.
- DePaulo, B. M., Stone, J. I., & Lassiter, G. D. (1985). Deceiving and detecting deceit. In B. Schlenker (Ed.), *The self and social life* (pp. 323-370). New York, NY: McGraw-Hill.
- deTurck, M. A., Feeley, T. H., & Roman, L. (1997). Vocal and visual cue training in behavioral lie detection. *Communication Research Reports*, 14, 249-259.
- Duchastel, P. (1989). ICAI Systems: Issues in Computer Tutoring. *Computers and Education*, 13, 95-100.
- Ekman, P. & O'Sullivan, M. (1991) Who can catch a liar? *American Psychologist*, 46, 913-920.
- Epstein, K., & Hillegeist, E. (1990). Intelligent Instructional Systems: Teachers and Computer-Based Intelligent Tutoring Systems. *Educational Technology*, 30, 13-19.
- Feeley, T. H. & deTurck, M. A. (1997). Case-relevant vs. case-irrelevant questioning in

- experimental lie detection. *Communication Reports*, 10, 35-45.
- Fisher, R. P., Brennan, K. H., & McCauley, M. R. (2002). The cognitive interview method to enhance eyewitness recall. In M. L. Eisen, J. A. Quas, & G. S. Goodman (Eds.), *Memory and suggestibility in the forensic interview* (pp. 265-286). Mahwah, NJ: Lawrence Erlbaum Associates.
- Fisher, R. P., & Geiselman, R. E. (1992). *Memory enhancing techniques for investigative interviewing: The cognitive interview*. Springfield, IL: Charles C. Thomas.
- Fisher, R. P., & Schreiber, N. (2007). Interview protocols for improving eyewitness memory. In: *The handbook of eyewitness psychology, Vol I: Memory for events*. Toglia, M. P. (Ed.); Read, J. D. (Ed.); Ross, D. F. (Ed.); Lindsay, R. C. L. (Ed.); Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, 53-80.
- Forrest, J. A., Feldman, R. S., & Tyler, J. M. (2004). When accurate beliefs lead to better lie detection. *Journal of Applied Social Psychology*, 34, 764-780.
- Frank, M. G., Feeley, T. H. (2003). To catch a liar: Challenges for research in lie detection training. *Journal of Applied Communication Research*, 31, 58-75.
- Freedman, R. (2000). What is an intelligent tutoring system? *Intelligence*, 11, 1-3.
- Gudjonsson, G. H. (2003). *The psychology of interrogations and confessions*. Chichester, England: Wiley & Sons.
- Gratch, J., Wang, N., Gerten, J., Fast, E., & Duffy, R. (2007). Creating Rapport with Virtual Agents. Proceedings of the 7th international conference on Intelligent Virtual Agents. Springer-Verlag, Paris, France, 125-138.
- Inbau, F. E., Reid, J. E., Buckley, J. P., & Jayne, B. C. (2001). *Criminal interrogation and confessions, 4<sup>th</sup> edition*. Gaithersburg, Maryland: Aspen.

- Johnson, M. K. (1988). Reality Monitoring: An experimental phenomenological approach. *Journal of Experimental Psychology: General*, 117, 390-394.
- Johnson, M. K., & Raye, C. L. (1998). False memories and confabulation. *Trends in Cognitive Sciences*, 2, 137-146.
- Kassin, S. & Fong, C. T. (1999). 'I'm innocent!' Effects of training on judgments of truth and deception in the interrogation room. *Law and Human Behavior*, 23(5), 499-516.
- Kassin, S. M., Meissner, C. A., & Norwick, R. J. (2005). "I'd know a false confession if I saw one": A comparative study of college students and police investigators. *Law & Human Behavior*, 29, 211-228.
- Köhnken, G. (1987). Training police officers to detect deceptive eyewitness statements: Does it work? *Social Behaviour*, 2, 1-17.
- Koedinger, K. R. & Corbett, A. (2006). Cognitive Tutors: Technology bringing learning science to the classroom. In Sawyer, K., *The Cambridge Handbook of the Learning Sciences*, Cambridge University Press, pp. 61-78
- Köhnken, G., Milne, R., Memon, A., & Bull, R. (1999). The cognitive interview: A meta-analysis. *Psychology, Crime, & Law, Special issue: The cognitive interview: Current research and applications*, 5, 3-27.
- Landry, K. L. & Brigham, J. C. (1992). The effect of training in criteria-based content analysis on the ability to detect deception in adults. *Law and Human Behavior*, 16(6), 663-676.
- Leo, R. A., & Ofshe, R. (1998). The consequences of false confessions: Deprivations of liberty and miscarriages of justice in the age of psychological interrogation. *Journal of Criminal Law and Criminology*, 88, 429-496.

- Lu, Z. L. & Eskew, R. T., (2007). A special issue on application of signal detection theory to visual perception. *Spatial Vision, 20, Special issue: Signal Detection Theory*, 1-4.
- Magenat-Thalmann, N. & Thalmann, D. (2004). *Handbook of Virtual Humans*. (p. 6-7). John Wiley and Sons.
- Mann, S. & Vrij, A. (2006). Police officers' judgments of veracity, tenseness, cognitive load and attempted behavioral control in real-life police interviews. *Psychology, Crime, & Law, 12*(3), 307-319.
- McArthur, D., Stasz, C., Zmuidzinas, M. (1990). Tutoring techniques in algebra. *Cognition and Instruction 7*(3) 197-244.
- Meissner, C. A., & Kassin, S. M. (2002). He's guilty!: Investigator bias in judgments of truth and deception. *Law and Human Behavior, 26*(5), 469-480.
- Mueller, S. T., & Weidemann, C. T. (2008). Decision noise: An explanation for observed violations of signal detection theory. *Psychonomic Bulletin & Review, 15*, 465-494.
- Olson, E. A., & Wells, G. L. (2004). What makes a good alibi? A proposed taxonomy. *Law and Human Behavior, 28*, 157-176.
- O'Sullivan, M., & Ekman, P. (2004). The wizards of deception detection. In Granhag, Pär-Anders and Strömwall, Leif (Ed.), *The detection of deception in forensic contexts* (pp. 269-286). New York, Cambridge University Press.
- Porter, S., McCabe, S., Woodworth, M., & Peace, K. A. (2007). Genius is 1% inspiration and 99% perspiration... or is it? An investigation of the impact of motivation and feedback on deception detection. *Legal and Criminological Psychology, 12*(2), 297-309.
- Porter, S., & ten Brinke, L. (2008). Reading between the lies: Identify concealed and

- falsified emotions in universal facial expressions. *Psychological Science*, 19, 508-514.
- Pspotka, J., Massey, L. D., & Mutter, S. A. (1988). *Intelligent tutoring systems: Lessons Learned*. Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc. pp. 552.
- Ruby, C. L., & Brigham, J. C. (1997). The usefulness of the criteria-based content analysis technique in distinguishing between truthful and fabricated allegations: A critical review. *Psychology, Public Policy, and Law*, 3, 705-737.
- Russell, T. A., Chu, E., & Phillips, M. L. (2006). A pilot study to investigate the effectiveness of emotion recognition remediation in schizophrenia using the micro-expression training tool. *British Journal of Clinical Psychology*, 45, 579-583.
- Santarcangelo, M., Cribbie, R. A., & Ebesu Hubbard, A. S. (2004). Improving accuracy of veracity judgment through cue training. *Perceptual and Motor Skills*, 98, 1039-1048.
- Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, 117, 34-50.
- 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 P. A. Granhag & L. A. Stromwall (Eds.), *Deception detection in forensic contexts* (pp. 64-102). Cambridge, England: Cambridge University Press.
- Sporer, S. L., & Bursch, S. E., (2003) Training to detect deception by verbal means: Learning to discriminate or change in response bias? Unpublished Manuscript, University of Giessen, Germany, 2003

- Sporer, S. L. & Schwandt, B. (2007). Moderators of nonverbal indicators of deception: A meta-analytic synthesis. *Psychology, Public Policy, and Law*, *13*, 1-34.
- Sporer, S. L. & Sharman, S. J. (2006). Should I believe this? Reality monitoring of accounts of self-experienced and invented recent and distant autobiographical events. *Applied Cognitive Psychology*, *20*(6), 837-854.
- Starns, J. J., Hicks, J. L., Brown, N. L., & Martin, B. A. (2008). Source memory for unrecognized items: Prediction from multivariate signal detection theory. *Memory & Cognition*, *36*, 1-8.
- Swartout, W., Gratch, J., Hill, R.W., Hovy, E., Marsella, S., Rickel, J., & Traum, D. (2006). Toward virtual humans. *AI Magazine*, *27*, 96-108.
- Verde, M. F., MacMillan, N. A., & Rotello, C. M. (2006). Measures of sensitivity based on a single hit rate and false alarm rate: The accuracy, precision, and robustness of  $d'$ ,  $A_z$ , and  $A'$ . *Perception & Psychophysics*, *68*, 643-654.
- Vrij, A., (2000). *Detecting lies and deceit: The psychology of lying and the implications for professional practice*. Chichester: Wiley and Sons.
- Vrij, A. (2005). Criteria-based content analysis: A qualitative review of the first 37 studies. *Psychology, Public Policy, and Law*, *11*, 3-41.
- Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities*. New York, NY, US: John Wiley & Sons.
- Vrij, A., Edward, K., & Bull, R. (2001). Police officers' ability to detect deceit: The benefit of indirect deception detection measures. *Legal and Criminological Psychology*, *6*(2), 185-196.
- Vrij, A., Edward, K., Roberts, K., & Bull, R. (2000). Detecting deceit via analysis of

- verbal and nonverbal behavior. *Journal of Nonverbal Behavior*, 24, 239-263.
- Vrij, A., Evans, H., Akehurst, L., & Mann, S. (2004). Rapid judgments in assessing verbal and nonverbal cues: Their potential for deception researchers and lie detection. *Applied Cognitive Psychology*, 18(3), 283-296.
- Vrij, A., Fisher, R., Mann, S., & Leal, S. (2006). Detecting deception by manipulating cognitive load. *Trends in Cognitive Sciences*, 10(4), 141-142.
- Vrij, A. & Graham, S. (1997). Individual differences between liars and the ability to detect lies. *Expert Evidence*, 5, 144-148.
- Vrij, A. & Mann, S. (2003). Telling and detecting true lies: Investigating and detecting the lies of murderers and thieves during police interviews. In M. Verhallen, G. Verhaeke, P. J. van Koppen, & J. Goethals (Eds.), *Much ado about crime: Chapters on psychology and law* (pp. 185-208). Brussel: Politeia.
- Vrij, A., Mann, S., Kristen, S., & Fisher, R. (2007). Cues to deception and ability to detect lies as a function of police interview styles. *Law & Human Behavior*, 31, 499-518.
- Vrij, A., Mann, S., & Fisher, R. (2006a). An empirical test of the Behavior Analysis Interview. *Law & Human Behavior*, 30, 329-345.
- Vrij, A., Mann, S., & Fisher, R. (2006b). Information-gathering vs. accusatory interview style: Individual differences in respondents' experiences. *Personality & Individual Differences*, 41, 589-599.
- Vrij, A., Mann, S. A., Fisher, R. P., Leal, S., Milne, R., & Bull, R. (2008). Increasing cognitive load to facilitate lie detection: The benefit of recalling an event in reverse order. *Law and Human Behavior*, 32(3), 253-265.
- Zuckerman, M., Depaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal

communication of deception. *Advances in Experimental Social Psychology*, 14, 1-59.

## FOOTNOTES

<sup>1</sup> Preliminary ANOVA's were conducted that included whether or not participants completed the interviewing task as a covariate. The effect of completing the interviewing task proved non-significant for both participants' discrimination accuracy ( $A_z$ ),  $F(1,101) = .18$ , *ns.*, response bias (C),  $F(1,101) = .19$ , *ns.*, or expressed confidence,  $F(1,101) = 1.77$ , *ns.*

**Table 1**

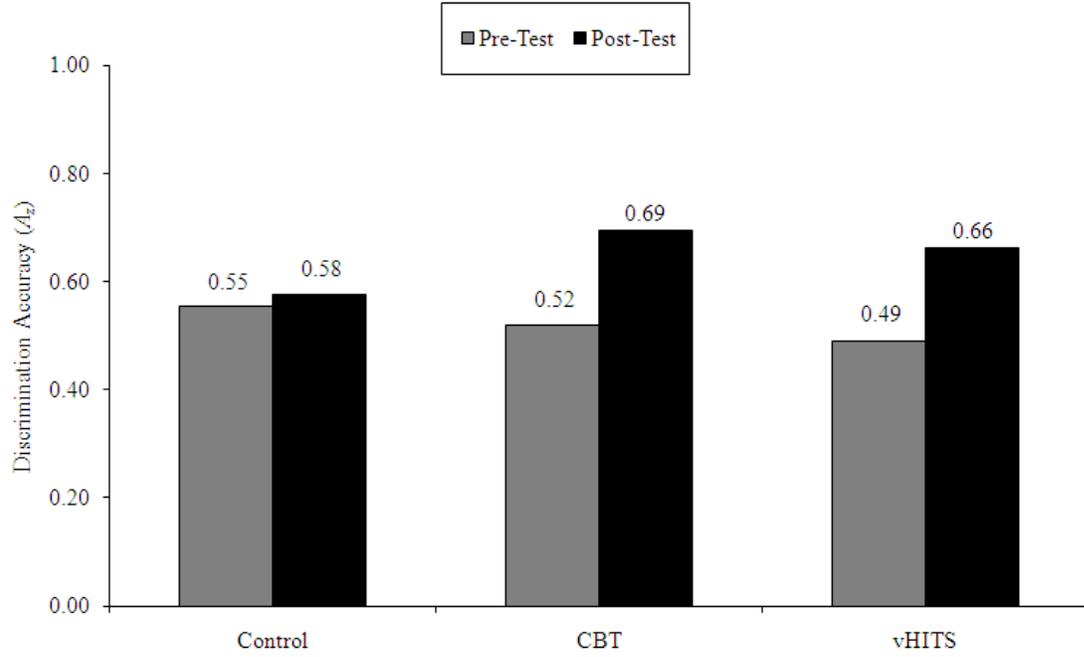
*Mean and standard deviation estimates for performance on the deception detection task across training conditions and pre-test/post-test*

	Control		CBT		vHITS	
	Pre-Test	Post-Test	Pre-Test	Post-Test	Pre-Test	Post-Test
Hits	0.63 (0.26)	0.74 (0.21)	0.55 (0.23)	0.82 (0.23)	0.62 (0.28)	0.85 (0.18)
False Alarms	0.54 (0.18)	0.61 (0.27)	0.53 (0.27)	0.68 (0.25)	0.64 (0.27)	0.71 (0.23)
Az	0.55 (0.23)	0.58 (0.27)	0.52 (0.24)	0.69 (0.19)	0.49 (0.29)	0.66 (0.24)
C	-0.30 (0.60)	-0.82 (0.98)	-0.05 (1.00)	-1.32 (1.26)	-0.54 (0.79)	-1.38 (1.09)
Confidence	83.97 (11.43)	85.74 (14.33)	80.89 (14.73)	84.80 (14.43)	84.09 (12.09)	91.83 (11.30)

**Table 2**

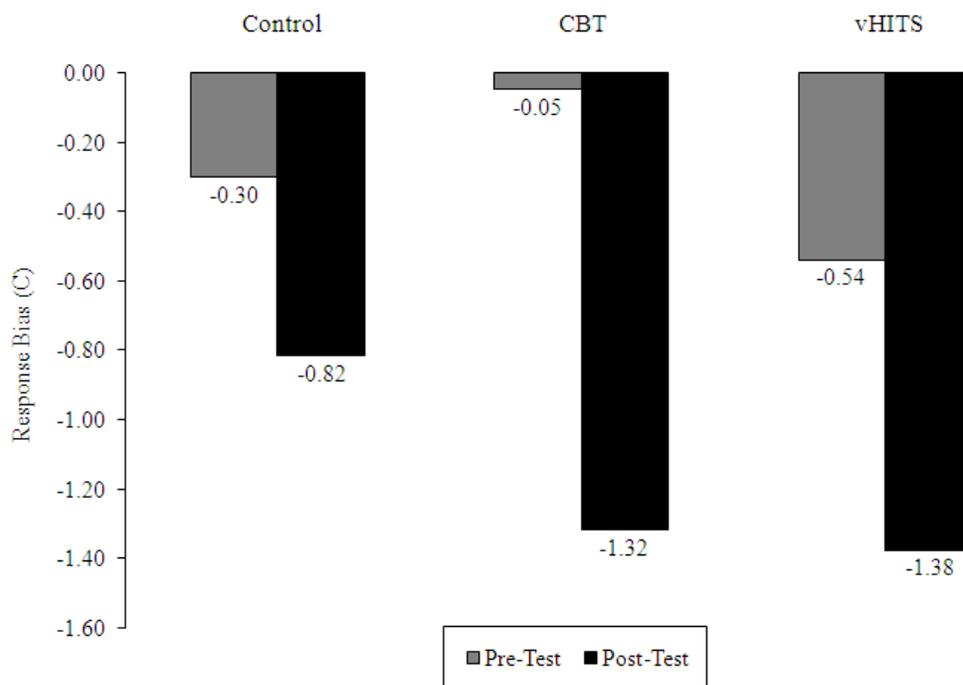
*Mean and standard deviation estimates for performance on the knowledge for cues to deception and interviewing skills questionnaire across training conditions and pre-test/post-test*

	Control		CBT		vHITS	
	Pre-Test	Post-Test	Pre-Test	Post-Test	Pre-Test	Post-Test
Verbal Cues	5.76 (1.36)	6.24 (1.63)	5.67 (1.77)	8.82 (1.33)	5.63 (1.50)	8.90 (1.62)
Nonverbal Cues	4.31 (2.09)	4.98 (2.24)	5.02 (2.30)	7.17 (2.69)	4.56 (2.21)	7.99 (2.82)
Interviewing Skills	6.30 (1.29)	6.70 (1.84)	6.63 (1.85)	9.19 (1.23)	6.81 (1.70)	9.82 (0.68)



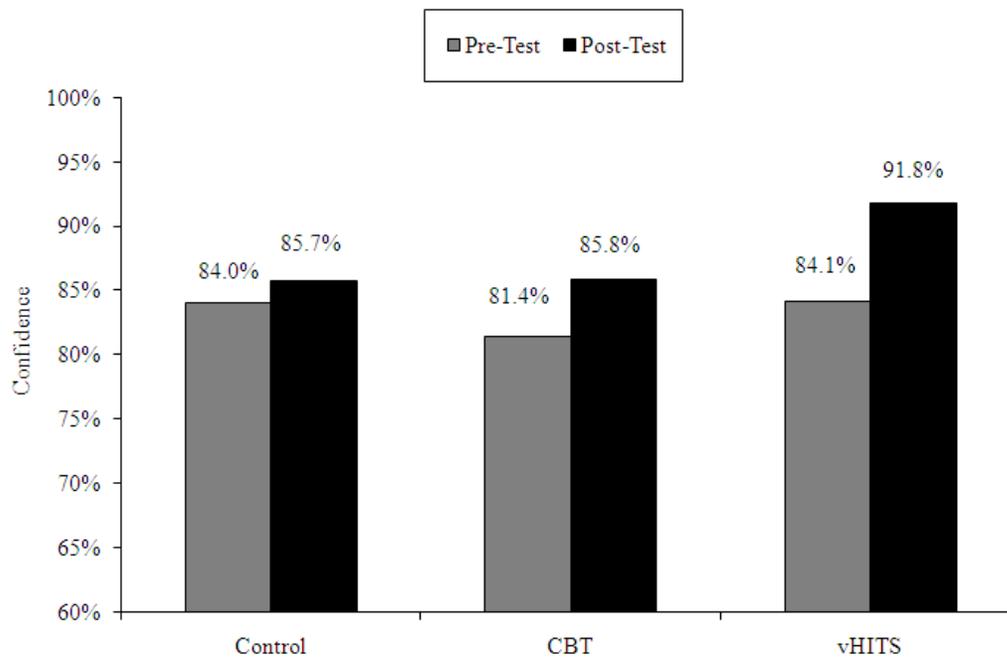
**Figure 1**

*Mean discrimination accuracy ( $A_2$ ) scores at both pre-test and post-test across all conditions*



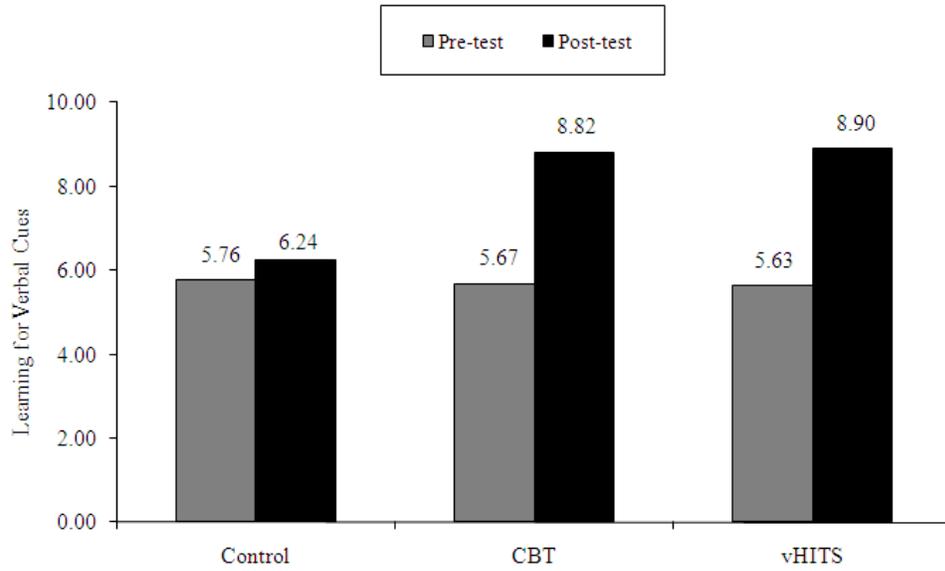
**Figure 2**

*Mean response bias (C) scores at both pre-test and post-test across all conditions*



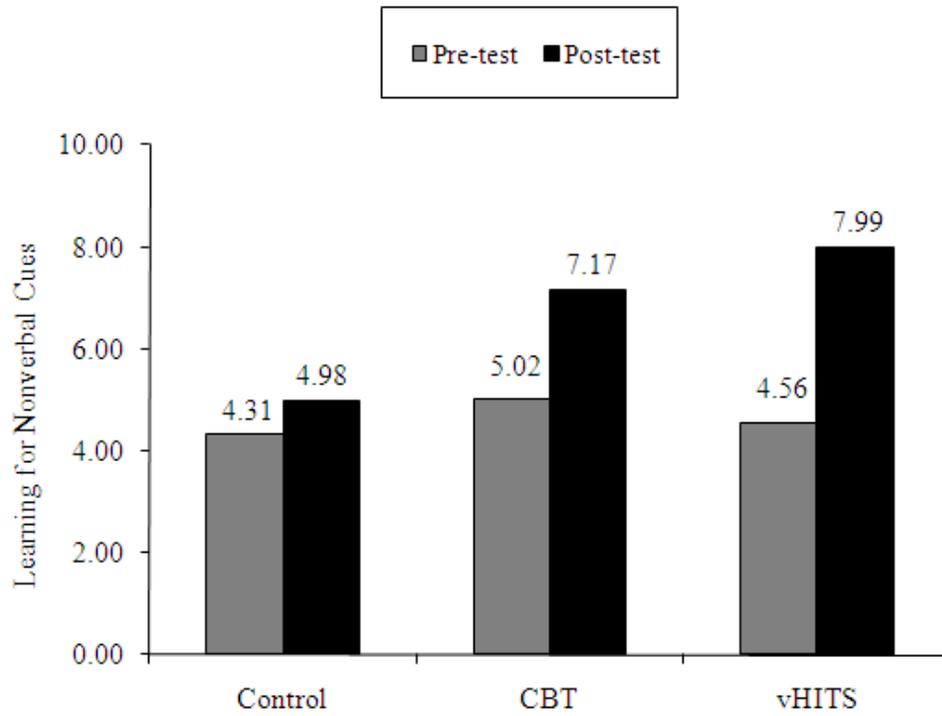
**Figure 3**

*Mean confidence scores at both pre-test and post-test across all conditions*



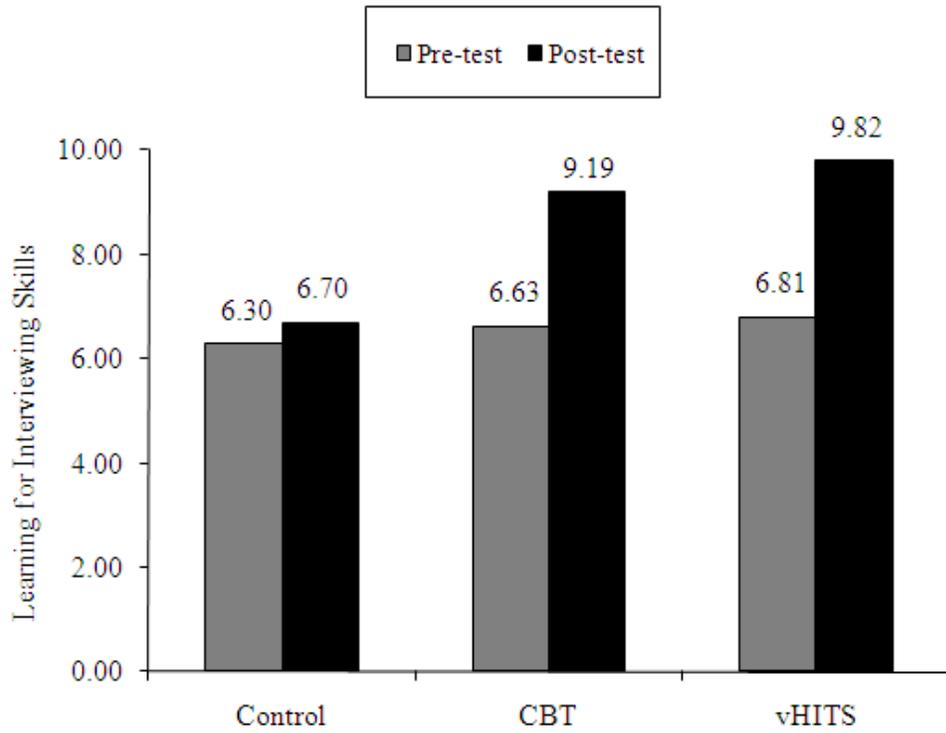
**Figure 4**

*Mean scores for knowledge of diagnostic verbal cues to deception at both pre-test and post-test across all conditions*



**Figure 5**

*Mean scores for knowledge of non-diagnostic nonverbal behaviors at both pre-test and post-test across all conditions*



**Figure 6**

*Mean scores for knowledge of investigative interviewing techniques at both pre-test and post-test across all conditions*

## Appendix A

### Questionnaire for knowledge of cues to deception and interviewing skills

#### Verbal Cues to Deception Questions

##### Instructions given to participant:

You will now answer several questions regarding verbal behaviors. Your task is to decide how often someone lying performs these behaviors vs. how often someone telling the truth performs these behaviors. Each verbal behavior will be identified and an example will be provided to you.

Then you will decide which number on the following scale best fits your view:

0     1     2     3     4     5     6     7     8     9     10

0 = Liars perform this behavior much more than truth tellers

5 = Liars and truth teller perform this behavior an equal amount

10 = Liars perform this behavior much less than truth tellers

##### Questions:

1. Logical structure means that a story is organized a reasonable way. For example, “First, I went out to eat. Then I went to the movies. After that, we went and got drinks.”
2. A spontaneous correction is when the right material is substituted for the wrong material without effort or premeditation. For example, “I was wearing a purple dress, and we went to the..., no wait, I was wearing my black dress.”
3. Admitted lack of memory is when someone acknowledges that they cannot recall certain details in a story. For example, “We went out to dinner, but I can’t remember what I ate.”
4. Details are small additions that provide more information in a story. For example, “I was watching the Red Sox game at Applebee’s while I drank a Miller Lite,” contains more details than “I was watching TV at a restaurant while I drank.”

5. Temporal and spatial details are small additions about what you see, hear, smell, touch, and taste that provide more information in a story. For example, “The whole place smelled like smoke and my drink tasted really sugary.”

6. Negative statements or complaints are comments that are not positive, express denial, and/or are uncooperative. For example, “I didn’t blow up the building and I don’t like being brought here against my will.”

### **Non-verbal Cues to Deception Questions**

#### **Instructions given to participant:**

You will now answer several questions regarding non-verbal behaviors. Your task is to decide how often someone lying performs these behaviors vs. how often someone telling the truth performs these behaviors. Each non-verbal behavior will be identified and an example will be provided to you. Then you will decide which number on the following scale best fits your view:

0    1    2    3    4    5    6    7    8    9    10

0 = Liars perform this behavior much more than truth tellers

5 = Liars and truth teller perform this behavior an equal amount

10 = Liars perform this behavior much less than truth tellers

#### **Questions:**

7. Gaze aversion is when a person refuses to make eye contact. For example, a person looks away from you when telling you a story.

8. Closed posture is when somebody’s body language seems guarded or tense. For example, a man may cross his arms when talking to you.

9. Tics are nervous behavior that people are unaware they are performing. For example a woman may rub her head, neck, or chin when she feels uncomfortable.

10. Gestures and signals made with your hands. For example, pointing at something is a gesture.

11. Hand and arm movements are motions made with your hands or arms. For example, a man might wave his hands in a circular motion to express exasperation.

12. Head movements are motions made with one's head. For example, a woman may nod her head for "yes" and shake her head for "no".

### **Accusatory Interviewing Questions**

#### **Instructions given to participant:**

You will now answer several questions regarding interviewing. Your task is to decide whether each of these behaviors is helpful or harmful when trying to detect lies.

Each interviewing behavior will be identified and an example will be provided to you. Then you will decide which number on the following scale best fits your view:

0      1      2      3      4      5      6      7      8      9      10

0 = This behavior is *not useful* when trying to detect lies

5 = This behavior is *neither useful nor un-useful* when trying to detect lies

10 = This behavior is *useful* when trying to detect lies

#### **Questions:**

13. A close-ended question is a question to which a person can only answer yes or no. For example, "Were you at the bar last night between the hours of 10 PM and 12 AM?"

14. An accusatory question contains a charge of wrongdoing. For example, "You killed your boss, didn't you?"

### **Investigative Interviewing Questions**

#### **Instructions given to participant:**

You will now answer several questions regarding interviewing. Your task is to decide whether each of these behaviors is helpful or harmful when trying to detect lies.

Each interviewing behavior will be identified and an example will be provided to you. Then you will decide which number on the following scale best fits your view:

0      1      2      3      4      5      6      7      8      9      10

0 = This behavior is *not useful* when trying to detect lies

5 = This behavior is *neither useful nor un-useful* when trying to detect lies

10 = This behavior is *useful* when trying to detect lies

**Questions:**

15. A rapport-building question builds trust and confidence between the interviewer and interviewee. For example, “Before beginning the interview, will you tell me a little bit about yourself?”

16. An open-ended question is a question which prompts an answer other than yes or no. For example, “What can you tell me about yourself?”

## **CURRICULUM VITA**

Justin Scott Albrechtsen was born on November 29, 1980 in Omaha, Nebraska. He is the son of Bryce and Dana Albrechtsen; the brother of Sadie Passmore, Jamie Albrechtsen, Nathan Albrechtsen, and Brent Albrechtsen; and the uncle of Ean Whitney, Aiden Whitney, Cole Whitney, Olive Passmore, Madeline Passmore, and Harrison Pardo. He graduated from Clinton High School in Clinton, Iowa in 1999. Upon graduation from high school, he attended Brigham Young University. In 2000 and 2001, Justin served a mission for The Church of Jesus Christ of Latter Day Saints, where he learned to speak Spanish fluently. Following his mission, Justin attended The University of Northern Iowa. Justin graduated from The University of Northern Iowa in May of 2005 with a Bachelor of Arts degree in Psychology and a Bachelor of Arts degree in Spanish. In August of 2005, Justin entered the Psychology Ph.D. program (Legal Psychology concentration) at the University of Texas at El Paso. Since entering the program he has worked under the supervision of Dr. Christian A. Meissner. Justin's research interests include understanding the social and cognitive processes involved in investigative interviews and deception detection. Justin has several research publications in the top-tier journals of Cognitive and Social Psychology and he has also given many conference presentations throughout The Unites States. In addition to his major research focus, Justin also has an interest in research methodology and statistical analyses that apply to predicting player and team performance in sports.

Permanent address: 544 Woodland Drive  
Clinton, IA 52732