Assessment Criteria Indicative of Deception (ACID): Replication and Gender Differences

Jessica A. Suckle-Nelson[1], Kevin Colwell[*,1], Cheryl Hiscock-Anisman[2], Samantha Florence[1], Kimberly E. Youschak[1] and Arianne Duarte[1]

[1]*Southern Connecticut State University, New Haven, CT, USA*

[2]*National University, La Jolla, CA, USA*

Abstract: Assessment Criteria Indicative of deception (ACID) is a statement analysis technique that combines careful interviewing designed to facilitate the detection of deception with empirically-derived content criteria related to both the phenomenal and objective nature of deception.. The current study of ACID had three goals. The first goal was to replicate ACID with a population of incarcerated witnesses. The second goal was to study the extension of this procedure by adding coherence and the type-token ratio for dependent measures. The third goal was to investigate the possibility of gender differences in verbal interpersonal deception. In meeting the first and second goals, this study demonstrated honest statements were longer and more detailed than deceptive statements, and that honest statements specifically had more words and more unique details added as a result of the recall enhancement effect of the interview. Honest statements were more coherent and had lower type-token ratios. In meeting the third goal, this study demonstrated that women's statements were more coherent and contained more words than male statements. Deceptive females gave less detail, fewer word, and had higher type-token ratios than deceptive men. Honest women gave more words and more detail than honest men. Importantly, coherence is very dependent upon gender, so that an honest man is more likely to tell an incoherent story than is a deceptive woman. Using gender-specific norms significantly increased the classification accuracy of a discriminant function analysis. Overall, 74 of 83 participants were accurately classified as honest or deceptive using the dependent measures. Future research is necessary to firmly establish the presence and cause of gender differences in verbal deception.

Keywords: Investigative interviewing, detecting deception, credibility assessment, gender differences.

INTRODUCTION

This paper replicated and extended an interviewing and assessment procedure referred to as Assessment Criteria Indicative of Deception (ACID; Colwell, Hiscock-Anisman, Memon, Taylor, & Prewett, 2008). Primarily, this study demonstrated the potential of ACID by drawing upon a group of participants especially relevant to the legal process: incarcerated men and women. Secondly, this paper evaluated the inclusion of additional content criteria to improve the ACID system. Finally, this paper identified several gender differences that were observed in the verbal deceptions of women versus men. The possible theoretical explanations for these differences will be discussed, as well as the practical implications stemming from their relationship to predictive accuracy.

ACID

Phenomenology of honesty versus deception. At the heart of this technique is the marriage of careful interviewing designed to facilitate the detection of deception with empirically-derived content criteria related to both the phenomenal and objective nature of deception. The Reality Interview (Colwell, Hiscock, & Memon, 2002[1]; Colwell et. al., 2008) is a permutation of the original Cognitive Interview that is structured specifically to highlight attempts at impression management and to increase the amount of cognitive effort expended by deceivers. The RI includes multiple recall attempts consisting of a free narrative and several mnemonic techniques that have demonstrated the ability to provide additional recall cues for honest people but to increase cognitive load for deceivers. These mnemonics include recall from another perspective, reverse order recall, and unanticipated questions in the form of forced-choice inferences, all of which serve to necessitate deeper processing of the original event (Colwell *et al.*, 2002; Colwell, Hiscock-Anisman, Memon, Rachel, & Colwell, 2007; Vrij *et al.*, 2008; Vrij *et al.*, 2009).

Honest respondents who are attempting to be convincing during an investigative interview believe that their honesty is transparent and that they will be believed as long as they do not make any major contradictions. Deceptive respondents also believe that they must avoid making major contradictions. However, in contrast to honest respondents, deceivers also are concerned with avoiding any verbal or nonverbal behaviors that could draw attention to them.

*Address correspondence to this author at the Department of Psychology, SCSU. 501 Crescent St., New Haven CT 06515, USA; Tel: +1 203 392 6871; Fax: +1 203 392 6805; E-mail: colwellk2@southernct.edu

[1] The RI was referred to as the Inferential Interview in Colwell *et al.*, 2002. The name was changed because it sometimes was mistaken for the intuitive interviewing that often is done by untrained professionals.

Significant to the present work, these beliefs include the need to avoid disclosing any piece of information that could lead to later detection. Thus, they try to avoid providing "too much" detail or adding "too many" details to their story after the initial free narrative. (Colwell, Hiscock-Anisman, Memon, Yaeger, & Milchlik, 2006; Hines *et al.*, 2009; Granhag, Andersson, Stromwall, & Hartwig, 2004). To accomplish these things, deceivers often create and practice a "lie script" and attempt to stay with this script as closely as possible during questioning. By doing so, they do not have to think about the original event and are, therefore, less likely to make a glaring contradiction or to release accidentally sensitive information.

In conclusion, honest respondents are free to think deeply about the original event with less effort and concern given to tracking the information they have presented. Honest respondents can disclose as much detail as they choose during the free narrative and are free to add as many additional details as they can remember as a result of the recall enhancement techniques during the mnemonic portion of the RI. Teherefore, honest respondents can take advantage of the spontaneous and automatic nature of genuine recall. In contrast, deceptive respondents must keep their lie script in mind and are not free to provide everything that they can remember during their free narrative. This means they must expend considerable cognitive effort to complete tasks such as reverse-order recall and unanticipated inferences. Additionally, the desire of deceivers not to add too much detail after their free narrative prevents them from benefiting as much as honest respondents from the recall enhancement effects of the mnemonics (Colwell *et al.*, 2006; Granhag *et al.*, 2004; Hines *et al.*, 2009; Vrij *et al.*, 2008).

Objective content criteria. ACID assesses patterns of speech and detail derived from the theories underlying Criterion-Based Content Analysis (CBCA), Reality Monitoring (RM), and impression management. These assessments include the response lengths and the numbers of external, contextual, and internal details as they are distributed across the free recall and mnemonics portions of the RI. Honest statements have been found to be more vivid and more spontaneous than deceptive statements. *Vivid* means that the statements are longer and more detailed, with more details coming from the senses and dealing with time and space. *Spontaneous* means that the statements are significantly longer after the free recall segment of the RI and contain significantly more unique details after free recall.

Extensions of the ACID technique. The initial research that created the RI also investigated deception using the content criteria of coherence and type-token ratio. A *coherent* statement is one that does not contain major contradictions and that could have happened given the limits of time and space (Colwell *et al.*, 2002). People typically know that it is important to tell a coherent story in order to be convincing and that sometimes an incoherent story is indicative of deception. However, this criterion, though predictive, has been left out of ACID research to date. Coherence was included in the present study due to the fact that people innately use this criterion anyway and because there is a potential difference between genders due to considerations that will be detailed below.

The *type-token ratio* (TTR) refers to the ratio of unique words in a statement to the number of total words in a statement. Thus, the statement, "One small step for man, one giant leap for mankind" has a TTR of 8 out of 10, or .80. TTR is believed to provide a measure of impression management. People who are working harder to communicate or who are trying harder to be convincing tend to use a wider array of vocabulary and, thus, have a higher TTR. This variable has shown the ability to predict honest from deceptive statements with inmates (Colwell *et al.*, 2002), US military personnel (Morgan, Colwell, & Hazlett, 2009), and Arabic suspects interviewed in English through an interpreter (Colwell, Hiscock-Anisman, Hazlett, & Morgan, 2009).

Several studies have used various combinations of these criteria, and the classification accuracies have ranged from 77% to near 95% (Ansarra *et al.*, 2009; Colwell *et al.*, 2009; Colwell *et al.*, 2002; Colwell *et al.*, 2007; Colwell *et al.*, 2008; Memon, Fraser, Colwell, Odino, & Mastroberardino, in press; Morgan *et al.*, 2009). Interestingly, it became apparent that those studies which focused on only one gender had classification rates near or above 90%, whereas those studies that combined genders into one sample had classification rates closer to 80%. Naturally, the question arose: Are there gender differences in verbal deception that may affect the system?

GENDER DIFFERENCES

Existing research has demonstrated significant gender differences in cognitive ability and interpersonal skill. Women often have significantly higher levels of verbal ability than men, though the effect sizes are typically small (Shibley Hyde & Linn, 1988; Weiss, Kemmler, Deisenhammer, Fleischhacker, & Delazer, 2003). It also has been reported that women have better attention to detail and better recall for episodic memory than do men (Duff & Hampson, 2001; Herlitz, Nilsson, & Backman, 1997; McBurney, Gaulin, Devineni, & Adams, 1997; Persinger & Richards, 1995; Postma, Jager, Kessels, Koppeschaar, & Van Honk, 2004).

Historically, women have been found to be higher in interpersonal and social sensitivity (e.g., nonverbal behavior) than are men, especially in situations where women have relatively less social power than their male counterparts (Crawford, 1995; Eagly, 1987; Henley, 1995; Lakoff, 1990). Related to this, research indicates that women may lie more frequently than men, and that women's lies more often are aimed at maintaining social cohesion rather than to benefit themselves (DePaulo & Kashy, 1998; DePaulo, Kashy, Kirkendol, Wyer, & Epstein, 1996; Feldman, Forrest, & Happ, 2002; Tyler & Feldman, 2004). This awareness of social cues in general, and the social nature of deception in particular, is likely mediated by the perceived social power as it manifests in learned gender roles (Eagly; Henley; Lakoff). Due to the observed differences in classification accuracy in previous ACID studies and the above research regarding women and deception, this study included gender as an independent variable.

THE CURRENT STUDY

The present study had three goals: (a) to replicate previous ACID findings with an incarcerated sample of

witnesses; (b) to investigate the extension of ACID to include the criteria of coherence and TTR; and (c) to assess for gender differences in verbal deception. For the first goal, it was hypothesized that honest statements would be longer and more detailed than deceptive statements. Further, it was predicted that honest statements would contain significantly more words and details during the free recall portion of the interview and would have more words and new details added in response to the mnemonics portion of the interview. For the second goal, it was hypothesized that honest statements would be more coherent and would have lower TTRs than deceptive statements. Finally, for the third goal, it was hypothesized that gender differences in verbal recall and deception would exist. The potential for gender differences was assessed in the form of: (a) main effects for gender in the dependent measures, (b) interactions between gender and honest in the dependent measures, and (c) classification accuracies of single-gender versus mixed-gender samples.

METHOD

Participants

Eighty-eight offenders (38 men, 50 women) from two facilities within the Texas Department of Criminal Justice-Institutional Division provided testimony regarding a staged theft. Their ages ranged from 18 to 57 years, with a mean of 28.3 years (*SD* = 3.9 years). The self-reported ethnicity of the sample was: 42% Caucasian, 36% African American, 18% Hispanic (primarily Mexican-American), 2% Asian, and 2% "Other". Administrative Segregation offenders were not included.

Procedure

The stimuli used in this study were identical to those used in the Colwell *et al*. (2002) and Colwell *et al*. (2007) studies but with new samples of offenders. Participants were told that they were in a "Memory Study" and were seated in classrooms of the Education Department of their respective prisons (one male unit and one female unit). They witnessed a staged theft, were given instructions about how to respond to the interviewer (honestly or deceptively), and then were interviewed by trained Research Assistants (RAs). For consistency, the same script was followed as closely as possible for each enactment of the theft, and the same actors and interviewers were used in the same roles at both units.

The staged theft was enacted at the front of the classroom. This event was shown to groups of inmates ranging from 8 to 17 members. No announcement marked the beginning of the scene, other than the arrival of new research staff in front of the classroom. This transition was quite obvious to the offenders and was augmented by the soon-to-be thief entering the room yelling (a behavior likely to draw attention in prison classrooms). The three central actors of the theft were seen by participants only during the enactment so that all participants had approximately the same exposure time in order to form memories. The entire simulation required approximately three minutes, and each participant had a clear view of the actors and materials[2].

After the portrayal of the theft, randomized instruction packets were distributed to all participants, denoting the honesty of reporting condition. Conditions were assigned in a manner so that any minor deviations in scene presentation were randomly distributed across experimental groups. Those in the honest condition were told to report everything they had witnessed as completely as possible. Those in the dishonest condition were told to distort their testimony in order to preclude conviction of the perpetrator. Specifically, dishonest respondents were asked to assume that they knew the perpetrator and to report in a manner that kept him out of trouble. Instructions were given in written as well as verbal form, and the same RA provided these instructions to all participants. Incentives could not be provided due to the potential for coercion with an incarcerated sample.

Following a short delay (ranging from approximately 5 to 20 minutes), participants were interviewed regarding the event. They were not allowed to discuss the event while waiting. Each participant was interviewed individually by a trained RA using a semi-scripted RI. Audio recordings of the interviews were transcribed and analyzed by blind raters. The transcripts from the male and female prisons were mixed together and could not be identified by raters.

Measures

This study investigated the traditional ACID content criteria with the addition of coherence and TTR. Central to the ACID system is the assessment of dependent measures at the free recall portion of the RI and then again across all of the mnemonic questions of the RI. For the traditional ACID criteria, this comprises the response length and the amount of external, contextual, and internal details presented during free recall of the RI, and the response length and the amount of new external, contextual, and internal details added across the mnemonic portion of the RI. The current study extends the traditional ACID criteria with the addition of coherence (measured once across the entire interview), and TTR. To keep the experiment-wise alpha level down, the number of dependent measures was reduced by summing the detail categories into totals for each of the two interview phases. Therefore, the study assessed coherence (Coh), free recall detail total (DetFR), free recall response length (ResFR), free recall TTR (TTRFR), mnemonics detail total (DetMN), mnemonics response length (ResMN), and mnemonics TTR (TTRMN).

Rater Training and Coding

Coh, DetFR, and DetMN are scored by hand, whereas ResFR, ResMN, TTRFR, and TTRMN are automated and scored by software. For the hand-scored data, the raters were trained according to the same procedures as those reported in Colwell *et al*. (2002) and Colwell *et al*. (2007). This training consisted of three one-hour group meetings with the second author, including homework assignments between two of the meetings and a final meeting to reach consensus. During the first meeting, the second author defined each of the three classes of detail and provided the raters with standard scoring sheets to ensure consistent operational definitions of the variables. He then described the procedure for tallying the amount and type of details present in a transcript. Raters were trained to tally each detail only the first time it appeared within a statement; that is, a detail that was

---

[2] The authors would like to thank John Yuille for providing us with this script.

mentioned in response to two different interview questions was tallied only in response to the first question. The result was that only the amount of novel detail elicited in response to each recall task of the interview was coded and analyzed. This allowed for the tracking of unique details as they were added throughout a statement, in order to highlight spontaneous additions.

Following the lecture portion of the training, raters practiced the detail tally procedure using an example transcript and then reviewed these codings with the second author. Raters scored a second practice transcript as homework and reviewed this as a group at the second training session. This process was repeated for a third transcript: Raters scored the transcript as homework and then reviewed this as a group at the third and final training session. At the final meeting, all discrepancies in codings were resolved and the raters engaged in one final round of discussion with the second author. Child witness statements from a previous training package (provided to the second author by John Yuille) were used as practice transcripts for training purposes so that raters would remain blind to the experimental conditions of the present study. All of these transcripts followed a precise script that was analogous to the interview script used in the present experiment.

Following these training sessions, each rater scored 10 more practice transcripts on their own. The interrater reliabilities, as measured by Pearson's *r* and average proportion agreement, were as follows: External = .91 (93%), Contextual = .86 (89%), Internal = .73 (78%). The most common difficulty was determining if a piece of information was irrelevant to the event in question and, therefore, an internal detail.

Each of the four raters rated half of the transcripts, meaning all transcripts were rated twice, in counterbalanced order. The values used for analysis were the average of the two raters' scores for each transcript. This was done to reduce rater bias due to individual differences or potential practice effects.

As per the ACID system, two values for each detail category were obtained: one for free recall, and one for the mnemonics. However, for the present study, the three detail categories were summed to yield total scores for each interview segment: the total amount of details provided during free recall (DetFR = External FR + Contextual FR + Internal FR), and the total amount of new details added during the mnemonics (DetMN = External MN + Contextual MN + Internal MN). The correlations and proportion agreements among the pairs of raters for the dependent variables was: DetFR, $r = .94$ (96%), and DetMN, $r = .88$ (90%).

Coherence (Coh) was defined as a story that did not contradict itself regarding major details or one that was possible given the limits of time and space. This variable was scored by hand, with each rater providing one score for an entire transcript. A transcript that did not contradict itself regarding major details or that contained a possible story was rated as *coherent* and received a score of 0. A transcript that did contradict itself regarding major details or that contained an impossible story was rated as *incoherent* and received a score of 1. This was done in order to make correlations and

effect sizes vary directly with contradictions. Given that each transcript was scored by two raters and these scores were averaged across raters, the actual potential scores for Coh were 0 (both rated coherent), .50 (one rated coherent and one rated incoherent), and 1 (both rated incoherent). The average proportion agreement between the paired raters' scores was .93 for Coh.

The automated data were scored by software written for the ACID system. As such, the test-retest and interrater reliability for ResFR, ResMN, TTRFR, and TTRMN all were 1.00. This was the same software package reported in Colwell *et al*. (2002) and Morgan *et al*. (2009).

RESULTS

Manipulation Checks and Experimenter Bias

All transcripts were read initially to make certain that each interview included the appropriate questions in the proper order and that participants followed instructions. One female and two male participants were dropped due to a failure to follow instructions. Multivariate outliers were investigated by comparing case numbers to Mahalanobis distances, resulting in an additional two cases being dropped. In all, five cases were dropped: 4 male and 1 female. This left 83 cases total: 34 male and 49 female. After dropping these cases, the data was normally distributed with homogenous variances.

Two one-way ANOVAs were run to assess for potential experimenter bias. The first examined the classification accuracy of each transcript by interviewer, $F(5, 76) = 0.83$, $p > .05$. The second examined the classification accuracy of by each group of raters, $F(2, 79) = 0.48$, $p > .05$. These non-significant findings indicate that there was no meaningful experimenter bias on behalf of interviewers or raters.

Mean Comparisons

The data first were analyzed *via* factorial MANOVA with the two factors of honesty (honest versus deceptive) and gender (female versus male). The dependent measures for this analysis were Coh, DetFR, ResFR, TTRFR, DetMN, ResMN, and TTRMN. Pillai's Trace indicated significant main effects for each factor and a significant interaction between the factors on the linear combination of the seven dependent measures: honesty, $F(7, 73) = 6.76$, $p < .05$, $\acute{\eta}^2 = .39$; gender, $F(7, 73) = 3.56$, $p < .05$, $\acute{\eta}^2 = .26$; and honesty by gender interaction, $F(7, 73) = 2.23$, $p < .05$, $\acute{\eta}^2 = .18$. The mean, standard deviation, and effect size for each dependent measure across honesty (gender collapsed) can be seen in Table **1**.

Honesty. Univariate ANOVAs were conducted for each of the seven dependent measures to assess further the observed effect for honesty. All seven of the dependent measures were significantly affected by honesty of reporting: Coh, $F(1, 79) = 6.06$, $p < .05$, SC = -.30; DetFR, $F(1, 79) = 42.16$, $p < .05$, SC = .92; ResFR, $F(1, 79) = 33.63$, $p < .05$, SC = .83; TTRFR, $F(1, 79) = 12.19$, $p < .05$, SC = -.52; DetMN, $F(1, 79) = 2.6$, $p < .05$, SC = .32; ResMN, $F(1, 79) = 5.92$, $p < .05$, SC = .32; and TTRMN, $F(1, 79) = 3.0$, $p < .05$, SC = -.30.

Gender. Similarly, univariate ANOVAs were conducted for each of the seven dependent measures to assess further

Table 1.   Mean Comparisons Across Honesty of Reporting (Total Sample)

Interview Stage Variable	Honest M (SD)	Dishonest M (SD)	F (1, 81)	p (One-Tailed)	SC Honesty
Total Interview					
Coherence*	0.06 (.24)	0.23 (.43)	5.00	.01	-.30
Free Recall					
# Details*	98.15 (47.45)	38.14 (23.90)	47.13	< .01	.92
# Words*	293.31 (161.70)	116.06 (62.55)	37.82	< .01	.83
TTR*	0.46 (0.13)	0.58 (0.15)	14.66	< .01	-.52
Mnemonics					
# Details*	42.88 (23.81)	28.77 (14.65)	2.33	.04	.32
# Words*	345.06 (214.78)	225.05 (140.89)	8.31	<.01	.39
TTR*	0.61(0.11)	0.67(0.13)	5.12	.01	-.30

Note: SC = Structure coefficient, a measure of effect size expressed as the correlation between the dependent measure and the overall observed differences due to honest versus deceptive responding. SC is interpreted in the same manner as a Pearson *r*.
* Denotes $p < .05$.

the observed effect for gender. The gender of the participant significantly affected four of the dependent measures: Coh $F(1, 79) = 10.86$, $p < .05$, $\acute{\eta}^2 = .12$; ResFR, $F(1, 79) = 3.14$, $p < .05$, $\acute{\eta}^2 = .04$; ResMN, $F(1, 79) = 9.82$, $p < .05$, $\acute{\eta}^2 = .11$; and TTRMN, $F(1, 79) = 5.66$, $p < .05$, $\acute{\eta}^2 = .07$. The gender of the participant did not significantly affect three of the dependent measures: TTRFR, $F(1, 79) = 0.01$, $p > .05$, $\acute{\eta}^2 < .01$; DetFR, $F(1, 79) = 0.62$, $p > .05$, $\acute{\eta}^2 < .01$; DetMN, $F(1, 79) = 0.02$, $p > .05$, $\acute{\eta}^2 < .01$.

Honesty by gender. The final set of univariate ANOVAs was conducted for each of the seven dependent measures to assess further the observed honesty by gender interaction. Results indicated a significant interaction for six of the dependent measures: Coh, $F(1, 79) = 4.79$, $p < .05$, $\acute{\eta}^2 = .07$; DetFR, $F(1, 79) = 2.86$, $p < .05$, $\acute{\eta}^2 = .03$; ResFR, $F(1, 79) = 3.70$, $p < .05$, $\acute{\eta}^2 = .05$; TTRFR, $F(1, 79) = 3.80$, $p < .05$, $\acute{\eta}^2 = .05$; ResMN, $F(1, 79) = 6.30$, $p < .05$, $\acute{\eta}^2 = .07$; and TTRMN, $F(1, 79) = 8.30$, $p < .05$, $\acute{\eta}^2 = .10$. Only one of the dependent measures was not affected by the combination of honesty and gender: DetMN, $F(1, 79) = 0.31$, $p > .05$, $\acute{\eta}^2 < .01$. The mean, standard deviation, and effect size for each dependent measure across each honesty and gender combination are reported in Table **2**.

Classification Accuracies

Three separate discriminant function analyses (DFAs) were used to classify statements as honest or deceptive, one for the combined sample and one for each gender independently. The purpose of using three DFAs was to assess the potential benefit of separating the data by gender in increasing the ability to detect deception. All three of the DFAs used the seven dependent measures to classify statements as honest or deceptive. The first DFA utilized the total sample (males and females), the second DFA utilized only male participants, and the third utilized only female participants. The results of each of these analyses are presented in Table **3**.

Table 2.   Mean Comparisons Across Honesty of Reporting and Gender

Female		
	Honest M (SD)	Dishonest M (SD)
Coherence	0.03 (0.18)$_{abc}$	0.05 (0.23)$_{abc}$
Free Recall		
# Details	105.51 (50.20)$_{ac}$	35.5 (23.5)$_{ac}$
# Words	332.58 (164.81) a,b,c	99.1(69.61)$_{abc}$
TTR	0.44 (0.13)$_{ac}$	0.61 (0.14)$_{ac}$
Mnemonics		
# Details	43.95 (29.82)$_a$	31.52 (21.64)$_a$
# Words	426.53 (213.03)$_{abc}$	236.30(143.32)$_{abc}$
TTR	0.56 (0.08)$_{abc}$	0.68 (0.14)$_{abc}$
Male		
	Honest M (SD)	Dishonest M (SD)
Coherence	0.11 (0.32)$_{abc}$	0.44 (0.52)$_{abc}$
Free Recall		
# Details	85.89 (40.83)$_{ac}$	41.25 (24.74)$_{ac}$
# Words	228.00 (136.89)$_{abc}$	118.38 (55.1)$_{abc}$
TTR	0.49 (0.13)$_{ac}$	0.55 (0.16)$_{ac}$
Mnemonics		
# Details	41.10 (40.56)$_a$	26.00 (18.41)$_a$
# Words	279.28 (137.42)$_{abc}$	211.69 (141.40)$_{abc}$
TTR	0.60 (0.11)$_{abc}$	0.66 (0.11)$_{abc}$

[a]Means differ significantly at the .05 level due to honesty of reporting.
[b]Means differ significantly at the .05 level due to gender.
[c]Means differ significantly at the .05 level due to honesty x gender interaction.

Table 3.   Classification Accuracies for ACID Variables (Gender-Combined and Gender-Split)

	Predicted Honest	Predicted Deceptive	% Correct
Gender-Combined			
True Honest	36*	12	75.0%
True Deceptive	5	30*	85.7%
Total		66/83	= 79.5%
Male Only			
True Honest – 18	16*	2	88.9%
True Deceptive – 16	2	14*	87.5%
Total – 34		30/34	= 88.25
Female Only			
True Honest – 18	26*	4	86.7%
True Deceptive- 16	1	18*	94.7%
Total		44/49	= 89.8%

Note: 79.5% of statements classified correctly for gender-combined; 89.1% of statements classified correctly for gender-split
*Denotes number of cases classified correctly.

Effect of gender on classification. A hierarchical linear regression assessed whether splitting the sample on the basis of gender led to a significant improvement in the DFA's ability to classify statements as honest or deceptive. Classification accuracies were computed for each case when predicted using the model with gender collapsed and for each case when predicted using the model that contained only their gender. The classification accuracy for each case computed using the gender-combined model was entered on the first block of the regression, and the classification accuracy for each case computed using the gender-split model was entered on the second block. This assessed whether the gender-split models led to an improvement in classification accuracy after controlling for the classification accuracy of the gender-combined model. Results indicated an improvement in classification accuracy using the gender-split models after controlling for the classification accuracy using the gender-combined model, $R_1 = .60$, $R_2 = .78$, $\Delta R = .18$, $\Delta F(1, 80) = 53.20$, $p < .05$.

DISCUSSION

Replication and Extension of ACID

The primary goal of this study was to replicate the ACID procedure with a sample of inmate witnesses. This was successful, as there was a significant relationship between the traditional ACID variables and honesty of reporting (i.e., honesty versus deception). As predicted, honest respondents provided longer responses and more overall detail than did deceptive respondents. Importantly, honest respondents also added more words and more details during the mnemonics portion of the RI. These findings are consistent with previous ACID research, where honest responding has been described as vivid (i.e., longer and more detailed) and spontaneous (i.e., more words and details added as a result of the recall

enhancement effects of the mnemonics; Colwell *et al.*, 2002; Colwell *et al.*, 2006; Colwell *et al.*, 2008).

The secondary goal was to extend the ACID system with the additions of coherence and TTR. This also was successful. Both honest and deceptive respondents made contradictions that resulted in incoherent statements. However, honest statements were more likely to be coherent than were deceptive statements. In previous studies (Colwell & Sjerven, 2005; Hartwig & Doering, 2009; Granhag *et al*, 2004; Hines *et al.*, 2009), both honest and deceptive respondents have indicated that it is necessary to avoid major contradictions in order to appear convincing. However, simply put, the additional effort required for successful deception meant that deceptive respondents were more likely to make glaring errors than were honest respondents, contributing to a significantly higher frequency of contradictions being made by deceptive respondents. These observed difference in coherence must be viewed with caution, however, as honest men were more likely to be incoherent than were deceptive women. In addition, honest statements demonstrated lower TTRs than did deceptive statements. In short, deceivers must work harder in order to be convincing than do honest respondents. This results in more careful use of language, which elevates the TTR. In sum, both additional criteria were related to deception in the predicted manner.

Classification accuracy. Consistent with previous research, approximately 80% (66 of 83) of participants in this study were classified accurately as honest or deceptive using a discriminant function analysis for the total (gender-combined) sample. This number increased to approximately 90% (74 of 83) accuracy when the sample was split by gender, such that a slightly different predictive model was used for each gender. This is consistent with observations in previous studies that showed relatively higher classification accuracies in single-gender samples.

Gender Differences

Gender. The third and final goal of this study was to explore gender differences in verbal deception. Statements provided by female participants were more coherent, contained more words during their initial description of the event, and contained more words during the mnemonics portion of the interview. Notably, deceptive women were even more coherent than honest men. These findings are consistent with previously reported gender differences in lexical ability and working memory (Duff & Hampson, 2001; Herlitz *et al.*, 1997; McBurney *et al.*, 1997; Postma *et al.*, 2004). The improved lexical ability of women leads to the potential to produce more words during a recall task. Hypothetically, superior working memory could make women less likely to contradict themselves as well, although more research is necessary before this can be stated with any certainty.

Gender by honesty interaction. Interestingly, it seems that the expected effects as a result of honesty versus deception per the ACID system were magnified by the effect of gender. That is, the performance of female participants on the variables dealing with details and response length are in the expected direction, but more polarized than the performance of male participants. Honest female participants

provided more words and more total detail at free recall and during the mnemonics phase than did both honest and deceptive male participants. However, deceptive female participants provided fewer words and less total detail at free recall and during the mnemonics than did all male participants. Finally, at least during free recall, TTR values indicated that deceptive women speak more carefully than any other group. The ACID system was created to pick out differences in speech that occur along cognitive and interpersonal lines. To the extent that culture teaches different cognitive and interpersonal skills to each gender, it makes sense that these gender differences should emerge during the ACID procedure. Women may lie more, though their reasons for doing so are indicative of interpersonal and social concern (DePaulo & Kashy, 1998; DePaulo *et al.*, 1996; Feldman *et al.*, 2002; Tyler & Feldman, 2004).

Historically, women have been found to be more aware of social and interpersonal cues (both verbal and nonverbal). Greater awareness of social cues and more experience with deception may combine to give women more knowledge about the process of interpersonal deception. However, this awareness is most likely mediated by the social power and status attributed to women rather than simply being a woman (Crawford, 1995; Eagly, 1987). Men typically have had more social power and status than women, and women who obtain more equivalent social power and status (e.g., by obtaining a higher position of authority within a company) have been found to communicate more like men. The United States incarcerates a much higher proportion of males than females, and those females who are incarcerated often are from positions of very low social status, due to histories of abuse, drug use, and prostitution. The groups in this study differed in gender and most likely also differed in social power.

FUTURE DIRECTIONS

The present findings suggest a number of future investigations in order to attempt to sort out the confluence of gender and social power. First, an attempt should be made to replicate the current findings in a new sample of male and female inmates. In addition, an attempt should be made to assess whether the findings occur in groups other than male and female inmates. Finally, future research also should study the relationship between social power and verbal behavior during interpersonal deception, both with and without the inclusion of gender as an additional factor. It is possible that the observed differences in this study are not really due to gender at all. They may be the result of the social conditions experienced by male versus female offenders.

CONCLUSIONS

The present study provided further evidence of the potential usefulness of the ACID procedure of interviewing and credibility assessment among a sample of forensically-relevant witnesses. Also, the study indicated that the inclusion of coherence and TTR as dependent measures may assist future credibility assessments. However, the low base-rate of violations of coherence and its confound with verbal ability and working memory suggest that this particular criterion should be used with extreme caution (especially among men). Finally, a number of gender differences were

seen, which indicate the need for future research in this area. Practically, these differences may mean that separate norms should be used for men and women when conducting verbal credibility assessments. Theoretically, future researchers should investigate the potential mediating effect that social status and perceived power may have on gender differences in verbal behavior during interpersonal deception. Even if these findings are not replicated in other social groups (e.g., because they are due to social power rather than to gender), the gender differences reported in this study may still be relevant to forensic professionals.

REFERENCES

Ansarra, R., Colwell, K., Hiscock-Anisman, C., Hines, A., Memon, A., & Kondor, S. (2009). Augmenting ACID with affective details to assess credibility. Manuscript under review.

Colwell, K., Hiscock-Anisman, C. K., Hazlett, G., & Morgan, C. A. (2009). Credibility assessment of suspects speaking Arabic through English interpreters. Manuscript in progress.

Colwell, K., Hiscock, C. K, & Memon, A. (2002). Interviewing techniques and the assessment of statement credibility. *Applied Cognitive Psychology, 16,* 287-300.

Colwell, K., Hiscock-Anisman, C., Memon, A., Rachel, A., & Colwell, L. (2007). Vividness and spontaneity of statement detail characteristics as predictors of witness credibility. *American Journal of Forensic Psychology, 25,* 5-30.

Colwell, K., Hiscock-Anisman, C. K., Memon, A., Taylor, L., & Prewett, J. (2008). Assessment Criteria Indicative of Deception (ACID): an integrated system of investigative interviewing and detecting deception. *Journal of Investigative Psychology and Offender Profiling, 4,* 167-180.

Colwell, K., Hiscock-Anisman, C., Memon, A., Yaeger, H., & Michlik, P. M. (2006). Strategies of impression management among deceivers and truth-tellers: How liars attempt to convince. *American Journal of Forensic Psychology, 24,* 31-38.

Colwell, K., & Sjerven, E. R. (2005). The 'Coin-in-Hand' stratagem for forensic assessment of malingering. *American Journal of Forensic Psychology, 23,* 83-89.

Crawford, M. (1995). *Talking difference: on gender and language.* Thousand Oaks, CA: Sage Publications, Inc.

DePaulo, B. M., & Kashy, D. A. (1998). Everyday lies in close and casual relationships. *Journal of Personality and Social Psychology, 74,* 63-79.

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). *Journal of Personality and Social Psychology, 70,* 979-995.

Duff, S. J., & Hampson, E. (2001). A sex difference on a novel spatial working memory task in humans. *Brain and Cognition, 47*(3), 470-493.

Eagly, A. H. (1987). *Sex Differences in Social Behavior: A Social-Role Interpretation.* Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Feldman, R. S., Forrest, James A., & Happ, B. R. (2002). Self-presentation and verbal deception: Do self-presenters lie more? *Basic and Applied Social Psychology, 24*, 163-170.

Granhag, P. A., Andersson, L. O., Stromwall, L. A., & Hartwig, M. (2004). Imprisoned knowledge: Criminals' beliefs about deception. *Legal and Criminological Psychology, 9,* 103-119.

Hartwig, M., & Doering, N. (March, 2009). Strategies of the deception game: A social cognitive theory of the psychology of innocence and guilt. Paper presented at the annual conference of the American Psychology-Law Society, San Antonio, Texas.

Henley, N. M. (1995). Body politics revisited: What do we know today? In Kalbfleisch, P. J. & Cody, M. J. (Eds.), *Gender, power, and communication in human relationships* (pp. 27-62). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Herlitz, A., Nilsson, L. G., & Backman, L. (1997). Gender differences in episodic memory. *Memory & Cognition. 25,* 801-811.

Hines, A., Colwell, K., Hiscock-Anisman, C., Ansarra, R., Memon, A., & Garrett, E. (Manuscript accepted pending revisions). Impression management strategies of deceivers and honest reporters in an investigative interview. *European Journal of Psychology Applied to the Legal Context..*

McBurney, D. H., Gaulin, S. J. C., Devineni, T., & Adams, C. (1997). Superior spatial memory of women: Stronger evidence for the gathering hypothesis. *Evolution & Human Behavior, 18*, 165-174.

Memon, A., Fraser, J., Colwell, K., Odino, G., & Mastroberardino, S. (in press). Distinguishing truthful from invented accounts using reality monitoring criteria. *Legal and Criminological Psychology.*

Morgan, C. A., Colwell, K., & Hazlett, G. (2009). Efficacy of forensic statement analysis in distinguishing truthful from deceptive eyewitness accounts of highly stressful events. Manuscript under review.

Persinger, M. A., & Richards, P. M. (1995). Women reconstruct more details than men for a complex five-minute narrative: Implications for right-hemispheric factors in the serial memory effect. *Perceptual and Motor Skills, 80*, 403-410.

Postma, A., Jager, G., Kessels, R. F. C., Koppeschaar, H. P. F., & Van Honk, J. (2004). Sex differences for selective forms of spatial memory. *Brain and Cognition, 54*, 24-34.

Shibley Hyde, J., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin, 104*, 53-69.

Tyler, J. M., & Feldman, R. S. (2004). Truth, lies, and self-presentation: How gender and anticipated future interaction relate to deceptive behavior. *Journal of Applied Social Psychology, 34*, 2602-2615.

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,* 253-265.

Vrij, A., Leal, S., Granhag, P. A., Fisher, R. P., Sperry, K., Hillman, J., *et al.* (2009). Outsmarting the liars: The benefit of asking unanticipated questions. *Law and Human Behavior, 33,* 159-166.

Weiss, E. M., Kemmler, G., Deisenhammer, E. A., Fleischhacker, W. W., & Delazer, M. (2003). Sex differences in cognitive functions. *Personality and Individual Differences, 35*, 863-875.