

**The Efficacy of CBCA and RM in Discriminating between
Reports of Single, Repeated, and Fabricated Events**

by

Jennifer A. A. Lavoie

B.A. (Hons.), Carleton University, 2001

**Thesis Submitted in Partial Fulfilment
of the Requirements for the Degree of
Master of Arts**

in the Department
of
Psychology

© Jennifer A. A. Lavoie 2004

SIMON FRASER UNIVERSITY

August 2004

All rights reserved.

This work may not be reproduced in whole or part, by photocopy or
other means, without permission of the author.

APPROVAL

Name: Jennifer Lavoie

Degree: Masters of Arts (Department of Psychology)

Title of Thesis: **The Efficacy Of CBCA and RM In Discriminating Between Reports Of Single, Repeated, and Fabricated Events.**

Chair: Dr. John McDonald
Assistant Professor, Psychology

Dr. Deborah Connolly
Senior Supervisor
Assistant Professor, Psychology

Dr. J. Don Read
Supervisor
Professor, Psychology

External Examiner: Dr. Nancy Hedberg
Associate Professor , Linguistics

Date Approved : August 6, 2004

SIMON FRASER UNIVERSITY



Partial Copyright Licence

The author, whose copyright is declared on the title page of this work, has granted to Simon Fraser University the right to lend this thesis, project or extended essay to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users.

The author has further agreed that permission for multiple copying of this work for scholarly purposes may be granted by either the author or the Dean of Graduate Studies.

It is understood that copying or publication of this work for financial gain shall not be allowed without the author's written permission.

The original Partial Copyright Licence attesting to these terms, and signed by this author, may be found in the original bound copy of this work, retained in the Simon Fraser University Archive.

Bennett Library
Simon Fraser University
Burnaby, BC, Canada

ABSTRACT

The present study examined the diagnostic capabilities of Criterion-Based Content Analysis (CBCA) and Reality Monitoring (RM) techniques in successfully distinguishing between reports of an event that were based on a single, repeated, or fabricated experience. Children (aged 7-8, $N = 60$) participated in a play session once, four times, or were coached to fabricate taking part in the event. One day after the target event, the children were interviewed for their memory of the play session. Recall reports were then coded using CBCA and RM criteria. Results indicated that initial global appraisals of perceived credibility failed to correctly assess authenticity of the account. Although single event reports were more likely to be classified as more credible than the other two conditions, accounts based on repeated experience were perceived only to be as credible as untrue accounts. Two models of Principal Component Analysis were generated in order to identify underlying dimensions of CBCA and RM, as well as to produce factor scores for further analyses. Derived components for the two credibility techniques were entered separately into two series of Multivariate Analyses of Variance (MANOVA). Individual and global evaluations of CBCA criteria did not discriminate real from fallacious accounts irrespective of event frequency. RM was sensitive to the veracity of statements, however the technique was unable to distinguish variability in event frequency. Implications regarding the perceived credibility assessments and the varying success of these credibility measures are discussed in relation to assessing verbal content of child witnesses in the criminal justice system.

DEDICATION

This thesis is dedicated to my parents, Robert and Anne.

ACKNOWLEDGEMENTS

First, I would like to thank Deb Connolly for her ongoing support from the very conception of this research project. Her encouragement, patience, and investment reviewing earlier drafts of this manuscript is appreciated more than could be expressed. Second, I would like to acknowledge all those who contributed to making this research possible: Jocelyn Conway and Jacquie Maloney for their skillful interviewing; Zina Lee and Jessie Klaver for their knowledgeable credibility coding, the multitude of teachers, principles and child care facilitators in the Greater Vancouver Region for their openness to the research process; and the RA's of the Connolly Lab for donating their time transcribing, and aiding in the completion of the finer details of this study. I would also like to thank Ray Koopman for his patience and availability for statistical advice and assistance. Thanks to Andrew, Sarah, Stephanie, Tristin, and Bryan for their support and friendship throughout this experience. This research was supported by a fellowship from the Natural Sciences and Engineering Research Council.

TABLE OF CONTENTS

Approval	ii
Abstract	iii
Dedication	iv
Acknowledgements	v
Table of Contents	vi
List of Tables	viii
List of Figures	ix
Introduction	1
Statement Validity Analysis (SVA)	1
Reality Monitoring (RM)	3
Children’s Reports for Instances of Repeated Events	5
Script theory	5
Fuzzy trace theory	6
Recent empirical evidence	7
Study Rationale	7
Method	9
Participants	9
Stimulus Event Procedure	9
Play sessions	9
Fabrication condition	10
Interviews	12
Rating of Transcripts	14
Personal credibility rating	14
CBCA	14
Total CBCA Rating	15
RM	15
Results	16
Descriptive Aspects and Analyses of Potential Artefacts	16
Incidence of the Ratings	18

Interrater Agreement22

Mean Ratings for CBCA and RM Criteria22

Inclusive Assessments: Personal Credibility and Total CBCA Ratings22

Principal Component Analyses of CBCA Criteria25

Principal Component Analyses of Reality Monitoring Criteria28

Discussion.....32

CBCA Approach.....33

RM Approach.....36

Perceived Credibility Rating.....38

Potential Limitations of the Study38

Conclusions and Future Research Directions39

References42

Appendix A: Letter of Ethical Approval48

Appendix B: Instructions for Fabrication51

Appendix C: Memory Interview Protocol51

Appendix D: CBCA and RM Coding Protocol.....54

LIST OF TABLES

Table 1:	Variable Details for Play Session Activities.....	11
Table 2:	Mean Number of Words, Sentences, and Average Word per Sentence by Condition.....	17
Table 3:	Incidence and Coefficients of Inter-rater Reliability for CBCA and Reality Monitoring Criteria for Two Raters.....	19
Table 4:	Means of CBCA and RM Criteria as a Function of Event Frequency.....	23
Table 5:	Factor Loadings of CBCA Criteria.....	26
Table 6:	Mean CBCA Factor Score and Standard Deviation (in parentheses) for Conditions of Event Frequency.....	27
Table 7:	Factor Loadings of RM Criteria	29
Table 8:	Event Frequency as a Function of Mean RM Component 2 Scores.....	30
Table 9:	Means and Standard Deviations (in parentheses) of Conditions of Event Frequency as a Function of CBCA Criteria and Total Assessment.....	35
Table 10:	Correlations of CBCA Criteria and RM Criteria.....	40

LIST OF FIGURES

Figure 1:	Perceived Credibility Rating of Single, Repeated and Fabricated Event Accounts.....	24
Figure 2:	RM Component 2 (External Cues) Means and Standard Error for Levels of Event Frequency.....	31

INTRODUCTION

An increase in the reporting and prosecution of child sexual abuse (CSA) in recent years has resulted in growing participation of children in the legal system (Bala, 1999). CSA cases are unique in that physical and medical evidence are often unavailable; thus, the child's statement is crucial in the investigation of suspected abuse (Lamb, Sternberg, Orbach, Hershkowitz, & Esplin, 1999). Child witnesses have been traditionally regarded by the Canadian legal system as inherently unreliable, possessing a proclivity for fantasy, and lacking in memory capability (Bala, 1999; Bala, Lee, Lindsay, & Talwar, 2000; Cashmore & Bussey, 1996). Bias in children's evidence is reinforced by the belief that fallacious reports produced by young clients are not rare (e.g., Porter & Yuille, 1996). The necessary dependency upon children's evidence coupled with the desire to detect spurious allegations, has generated a production of psychological innovation designed to address the reliability of children's evidence.

Psychological scholarship dedicated to the detection of deception has yielded four general research approaches: personality characteristics, non-verbal behaviour, physiological measurements, and content-oriented statement assessment (Sporer, 1997). The latter approach consists of two general lines of research established to discriminate between truthful and deceitful accounts based on systematic differences in linguistic structure and content: Statement Validity Analysis (SVA) and Reality Monitoring (RM) (Porter, Yuille, & Lehman, 1999; Sporer, 1997).

Statement Validity Analysis (SVA)

The most commonly used technique for assessing children's verbal evidence is SVA. Originally developed in Germany, SVA was formulated to evaluate the validity of children's allegations of sexual abuse. The SVA technique is based on the *Undeutsch hypothesis*, the notion that statements based on self-experience differ in content and structure from

statements invented from fiction (Undeutsch, 1989). Central to this hypothesis is the belief that certain information is likely to be present in a statement elicited by a person who has truly experienced an event but not in a report based on fiction. SVA consists of three components: (a) a structured interview to obtain a complete statement, (b) a systematic analysis of verbal content (Criterion-Based Content Analysis), (c) and the Validity Checklist, an evaluation of the outcome of CBCA and the testing of alternative hypotheses (Raskin & Esplin, 1991; Steller & Köhnken, 1989). Criterion-Based Content Analysis (CBCA) is the central component of SVA and consists of a set of 19 criteria. The statement is rated for the presence or absence of these criteria, where higher incidence of these criteria reflects greater credibility of the account. These criteria are subsumed under five subscales: (a) General Characteristics (e.g., logical structure, quantity of details), (b) Specific Content (e.g., reproduction of conversation), (c) Peculiarities of Content (e.g., unusual details), (d) Motivation-Related Content (e.g., pardoning the perpetrator), (e) Details Characteristic of the Offence which are based on criminal findings (Raskin & Esplin, 1991; Steller & Köhnken, 1989).

The outcome of studies empirically validating the CBCA approach are mixed (Vrij, in press). In the field, validation studies have revealed that this model discriminates reports of real events from those that are false (Ruby & Brigham, 1997). In the laboratory, the detection of deception using CBCA is not as successful. Porter and Yuille (1996) found that number of details reported, coherence, and admissions of lack of memory were the only criteria that significantly differentiated adult's truthful and deceptive accounts. In examining adults who were asked to recall real and fabricated childhood events, Porter and Yuille (1996) found that reporting a lack of memory was indicative of real memories and repeated details were associated with fabricated accounts. In a study investigating children's accounts of experienced and falsified accounts of a mildly traumatic event, Santtila, Roppola, Runtti, and

Niemi (2000) found that age and verbal ability increased the occurrence of certain CBCA criteria regardless of truthfulness and that various criteria differentiated between accounts within specific age groups on age group (7-8, 10-11 and 13-14 years). *Logical structure* and *quantity of details* were the only two criteria that discriminated between true and false accounts for all age groups. Sporer (1997) examined the validity of the first 13 CBCA criteria in distinguishing between fabricated and self-experienced adult accounts. He found that true accounts consisted of significantly more *logical consistency* and *contextual embedding* than fabricated accounts. All in all, despite the equivocal diagnostic ability of CBCA, the SVA technique is currently being used in legal settings in Germany and other parts of Europe (Vrij, Akehurst, Soukara, & Bull, 2002).

Reality Monitoring (RM)

A second approach formulated to evaluate the credibility of statements is RM, originally proposed by Johnson and Raye (1981). The process of RM refers to distinguishing memories of internally derived information (e.g., thoughts, imaginings) from memories based on perceived external information, such as texture and feelings (Johnson, Hastroudi, & Lindsay, 1993). RM is a particular form of Source Monitoring, which is the process of identifying the particular origin of a memory by making attributions based on information available in activated memories (Johnson et al., 1993). Contrary to the notion that the source information is simply contained in a tag, Johnson and Raye asserted that source is determined through an assessment of attributions based on information activated with the knowledge being retrieved. At encoding, there is information that can be used to make source attributions (e.g., voice, kinetic information, visual details). The same information that is available at perception that allows one to know the experience is real is presumably available to make the same source attributions at recall. An assumption of RM is that memories of real events are more likely to contain more perceptual information (i.e., visual and auditory details, taste,

touch and smell), semantic information (i.e., meanings and connections), contextual information (i.e., details relating to time and place), and affective information (i.e., feelings and emotions during event). Alternatively, memories derived internally from imagination are expected to contain more subjective details and cognitive operations (thoughts, and reasoning) (Johnson et al., 1993; Johnson & Raye, 1981).

Preliminary research suggests that criteria based on RM are effective in discriminating between true and false statements. Studies applying the RM approach found that truthful reports stated immediately after the event, contained more sensory and contextual information than fabricated statements (Alonso-Quecuty, 1992, 1996; Hernandez-Fernaud & Campos, 1997; Johnson, Foley, Suengas, & Raye, 1988; Sporer, 1997; Strömwall, Bengtsson, Leander, & Granhag, in press). The RM approach has been formalized in a number of ways; first by Johnson et al. (1988) who developed the Memory Characteristics Questionnaire (MCQ), and later by Sporer and Kupper (Judgment of Memory Characteristics Questionnaire, 1995) and Sporer (1997).

While studies examining the empirical validity of CBCA and RM directly suggest that both CBCA and RM are comparatively effective in identifying truthful and fabricated statements (Sporer, 1997; Vrij, Edward, Roberts, & Bull, 2000), this validation has primarily been based on comparisons of the reports of a false event and an event occurring a single time (see Strömwall et al., in press, for a notable exception). Given that this research is extended to children who experience abuse, and that children generally experience multiple instances of victimization, it is unclear if these measures are suitable for evaluating the truthfulness of an instance of a repeated event versus a fabricated event. Theoretical and empirical evidence suggest that memory for unique and routine events are distinctive.

Children's Reports for Instances of Repeated Events

Given the relation between cognition and verbal expression, the analysis of language has been used to make inferences about underlying cognitive representations (e.g., Nelson, 1986; French, 1986). The distinction between the verbal report and the underlying memory representation is an important one; although knowledge that is recalled is part of the representation, not all information that is represented is necessarily recalled (Fivush, 1997; Nelson, 1986). Restricted recall may be tied to the reason for recalling the information, as well as to developmentally related limited verbal expression. Despite the rather tenuous direct link between verbal recall and contents of memory, several theorists maintain that verbal accounts associated with various types of events (i.e., unique, repeated experiences) differ qualitatively.

Script theory. According to Script theory, when experience with an event increases, an abstract, schematically organized, representation of what typically occurs during the routine event (i.e., script) develops in memory (Nelson, 1986; Shank & Abelson, 1977). These scripts are spatially and temporally organized sequences of actions, actors, and objects likely to be present during any given occurrence of the event. When children recount familiar events, they typically report the component actions in correct temporal sequence where logically ordered events are recounted in the appropriate order, and arbitrarily ordered events are more flexibly recounted. In fact, children as young as 3 report recurring events in ways qualitatively similar to adults (Fivush, 1997).

Because scripts represent what generally happens during any given occurrence of an event, the focus of what usually happens leads to a potential loss of information about specific instances of an atypical event. Consequently, the description of an instance of a recurring event becomes generalized and diminished in distinctive details (Shank & Abelson, 1977). Fivush (1997) suggests that children report recurring events more generally in that

children do not report a single instance of the event, rather, they report what typically happens every time the event occurs. Further, these events are recounted by children in the timeless present tense (e.g., “run”, “eat”) and in the second person, “you”, suggesting that they are recounting the actions of a person experiencing the situation, and not their own personal experience. Additionally, compared to a report of a specific event, increased use of temporal markers (e.g., “then”, “next”), and definite articles (e.g., “the” to introduce nouns not previously mentioned) were observed in narratives of children who had greater familiarity with the event in question (Fivush & Slackman, 1986; French, 1986; Nelson & Gruendel, 1986).

With increasing experience with events, children report more component actions and more conditional action (if, then sequences), though older children report more actions, more complexity even with equal amount of experience. Children’s event representations become more complex with age and experience with specific events (Fivush & Slackman, 1986). That is, they report more alternatives (e.g., play with the doll *or* the blocks), and more optional activities (e.g., you *might* get to play with the bike). In a study involving unique and repeated experiences of kindergarteners, Fivush and Slackman found that increased experience with an event gave rise to accounts that were more schematic, temporally complex, and hierarchically organized. Increasing complexity of script reports with age is due to developing language skills, as well as increasing ability to represent events in more complex ways (Fivush, 1997).

Fuzzy trace theory. Fuzzy trace theory has also been used to distinguish between reports varying in event frequency. According to Brainerd and Reyna (1998), gist (general details) and verbatim (specific details) memory of events are stored independently, and verbatim traces decay more quickly than gist memory such that memory becomes more gist-based over time. Memory for true events contains both gist and verbatim information while memory for fabricated events contain relatively fewer verbatim details. Hence, recall of a true

experience should contain more specific details; however, when recalling an instance of a repeated event, the forgetting and recovery of these verbatim details may lead to inconsistency of truthful accounts over time relative to accounts of fabricated events.

Recent empirical evidence. Memory reports of experienced events have recently been demonstrated to differ qualitatively based on whether the event to be recalled was experienced once or repeatedly (Connolly & Lindsay, 2001; Powell & Thompson, 1996, 1997, 1999, 2003). These studies share a protocol where children participate in an event that contains a series of activities once or multiple times. Children participating in repeated occasions are exposed to a number of fixed (i.e., details that are experienced the same way during each instance) and variable details (i.e., details that are experienced differently during each instance) across instances. In comparing children's reports of the final occasion with those reports based on a single experience, it has been found that increased experience lead to reporting more errors about the event (Connolly & Lindsay, 2001; Price & Connolly, in press). In contrast, Powell and Thompson (1997) demonstrated that repeated experience with similar occasions strengthens the event memory.¹ Thus, given the theoretical and empirical evidence indicating that single and repeated experiences are processed and expressed differently, the efficacy of statement criterion analysis in appropriately classifying truthful and deceitful accounts regardless of event frequency is questioned.

Study Rationale

This study was designed to investigate children's verbal accounts of self-experienced unique events, instances of a repeated event, and fabricated events to determine whether elements of truthful accounts can be reliably differentiated from fallacious accounts through the use of established credibility criteria. At present, there has been no systematic comparison of memory reports of fabricated, unique and instances of repeated events. The

¹ This discrepancy is likely due to differences in the degree of similarity between variable detail options between the studies.

issue of memory for repeated experience is important because most children who disclose allegations of sexual abuse report multiple instances of victimization (Connolly & Read, 2003). The rationale of this study was to examine whether existing credibility assessment protocols (specifically, CBCA and RM) were sensitive to the veridicality of children's reports irrespective of event frequency. Given the demonstrated differences in memory recall for experienced events based on event frequency, it was hypothesized that these credibility measures would not discriminate reports of an instance of a recurring event from a false event report as successfully as reports of single and fabricated events. It is important to discover the efficacy of these measures on reports of varying event frequency in a controlled environment in order to inform future use of these tools.

METHOD

Participants

Seventy-three children were assigned to one of three conditions: single event (SE), repeated event (RE) or a event fabrication condition (FE). Eleven participants did not complete the study. Reasons for attrition included absence due to illness ($n = 5$) or early departure from school ($n = 6$). A mechanical failure in the recording device also prevented interviews in two cases from being transcribed. Sixty children (33 female and 27 male) aged 7 to 8 ($M = 7.43$ years; $SD = 0.50$ years) recruited from local schools and after-school-care facilities in the Greater Vancouver area completed the study. Written parental permission and individual permission was obtained from each child to participate as well as to be audio taped during the final memory interview. Ethical permission to conduct this study was obtained from the Simon Fraser University Ethics Committee (see Appendix A).

Stimulus Event Procedure

Play sessions. Participants were invited to participate in a 15-minute play session that took place at the child's school or care facility. For the children in the SE and RE conditions, play sessions were conducted individually with each child. During each play session, the participant was given materials, instructions, and assistance to engage in the following activities: a warm-up exercise, a 15-piece puzzle, a magic trick, and a science game. The order of the activities was fixed.

In the RE condition, play sessions occurred over two consecutive days, with separate sessions occurring in the morning and in the afternoon. The target play session (i.e., the play session to be recalled) was the single play session for children in the SE condition, and the last play session on day two for children in the RE condition. During the target play session, the experimenter wore a red apron and described the event as "Apron Day" in order to tag the play session for subsequent recall.

As illustrated in Table 1, each play session consisted of four activities and 10 critical details. The details associated with each of the four play sessions are listed in Table 1. The last session listed in each order grouping (# 4) was the target event to be remembered. The sample was partially counterbalanced across sessions where there were two random orders of variable details in the RE and SE condition, and four random orders of variable details in the SE condition. Each of the first four rows listed under RE Order 1 represent the four randomly ordered target events for the SE condition. Session 3 and 4 under RE Order 1 each represent the two randomly ordered target events for the FE condition.

The play sessions proceeded as follows: First, the children were asked to do one of four warm-up exercises (e.g., jumping jacks). Second, children listened to music played by one of four instruments while they constructed one of four picture puzzles and wore one of four occupational hats. Next, children were given a magic aid (e.g., a wand), and instructions to assist in the performance of one of four tricks while reciting one of four magic phrases. Finally, children carried out one of four science games where the children placed one of four stickers on their hands, and used one of four materials smelling of various fruit. The children were informed that they would be interviewed the following day.

Fabrication condition. For the fabrication condition, a coaching paradigm was chosen because it was felt that while embellishments of a child are easily identifiable, accounts developed in conjunction with the coaching of an adult are of more concern because they are more difficult to detect. For the children in the FE condition, the session was conducted individually with each child and lasted 15 minutes. Participants were told that the experimenter was interested in their story telling abilities. Children were presented a series of twelve 8" x 11" black and white line drawings depicting each of the activities in the target event (e.g., a picture of a police hat worn by a child constructing a puzzle). The storyboard of a novel event was chosen as a medium for coaching to overcome the confound of

Table 1.

Variables Details for Play Session Activities

<u>Warm-up</u>	<u>Puzzle Activity</u>			<u>Magic Activity</u>			<u>Science Activity</u>		
Exercise	Picture	Music	Hat	Trick	Words	Aid	Project	Smell	Sticker
<u>RE ORDER 1</u>									
1. Stretching	Elephant	Violin	Construction	Pencil	Alacazam	Ring	Weather	Apple	Boat
2. Toe Touching	Lion	Guitar	Explorer	Ball	Abracadabera	Glove	Planet	Coconut	Car
3. Running	Tiger	Piano	Firefighter	Cube	Hocus Pocus	Wand	Flower	Lemon	Plane
4. Jumping Jacks	Zebra	Drum	Police	Penny	Shazam	Scarf	Fossil	Blueberry	Train
<u>RE ORDER 2</u>									
1. Jumping Jacks	Lion	Drums	Firefighter	Pencil	Abracadabera	Scarf	Planets	Coconut	Plane
2. Stretching	Zebra	Violin	Explorer	Cube	Alacazam	Glove	Fossils	Berry	Boat
3. Toe touching	Tiger	Piano	Police	Ball	Shazam	Ring	Flowers	Lemon	Train
4. Running	Elephant	Guitar	Construction	Penny	Hocus Pocus	Wand	Weather	Apple	Car

constructing the fabrication from a pre-existing experience, as well as to control the content of the fabrication to make a direct comparison with the experienced event. It was felt that this medium would be generalizable to the kinds of sources that children use to facilitate a fabrication (e.g., information from another person, book, television). The storyboard pictures were introduced as describing events that occurred on 'Apron Day.'

The children were coached to generate a convincing story about themselves and the experimenter participating in the activities depicted in the pictures. The experimenter prompted the children to view the picture series twice while actively constructing a story. Misinterpretations of critical details were clarified (i.e., the child misunderstood a drawing). Participants were asked to attempt to convince the interviewer, who would speak with them the next day, that the story events had actually occurred (see Appendix B for procedural instructions for the FE condition). The target event for children in the fabricated condition was the single session in which they were asked to construct a story using pictures. The order of variable details was (partially) counterbalanced. The children were informed that they would be interviewed the following day and they were instructed to try to convince the interviewer that the activities depicted in the pictures had really happened. The initial session and the memory interviews was audio taped.

Interviews

Memory interviews were individually administered to each participant one day following the target event by one of two new interviewers (both female). Interviewers were told that the interviews were part of a study on memory, but were blind to the research design and hypotheses. Interviewers were given a brief training session relating to general interviewing techniques including skills such as establishing rapport, avoiding leading questions, avoiding interruptions and permitting pauses (see Poole & Lindsay, 1995; Raskin & Esplin, 1991; Yuille, 1988, for recommended interviewing guidelines).

Interviewers were given a protocol guide to facilitate the structuring of the interview and to maintain as much similarity between interviews as possible. After establishing rapport, the purpose of the interview was described to the child and the participant's were asked if they remembered the target event. Free recall was then initiated by asking the participants to recount as much as possible about the target event: "Can you tell me everything you can remember about 'Apron Day'?" (See Appendix C for interview protocol and questions).

Because open-ended questions reliably yield more accurate responses relative to specific prompts (Orbach & Lamb, 2000; Saywitz & Camparo, 1998), most of the interview was focused on gathering information from responses to open-ended prompts. Following the initial free recall question, further retrieval attempts were made by asking the participants if they could remember more. This was accomplished through administering five scripted free recall prompts concerning memory for the target event (e.g., Do you remember anything else? Even things you don't think are important?). A final retrieval attempt was made by asking participants if they could remember additional information about selected issues that the participant had previously raised ("You mentioned that you _____. Can you tell me anything more about that?"). Given the unique obstacles that arise in interviewing children, this semi-structured process allowed some flexibility for the interviewer to question the child in a fashion that was personally suited to the child such that more clear and meaningful information could be obtained.

After the child expressed that she could not recall any more about the event, the interviewer administered a cued recall test where the child was asked to respond to specific questions about each activity (e.g., "You put together a puzzle during apron day, what picture was on the puzzle?"). These cued recall data were not used for this study, but were collected for future analyses. At the conclusion of the interview, the participants were praised, thanked for their help, and asked if they had any questions. There was no time limit placed on the

duration of the interview. All interviews were audio taped and subsequently transcribed verbatim.

Rating of Transcripts

All transcripts were coded by two raters (both female) who were trained in both Criterion-Based Content Analysis (CBCA) and the Reality Monitoring (RM). Each rater received and was instructed to read a training manual consisting of detailed explanations of each of the reality criteria from both scales. A training session was then conducted where each criterion was discussed thoroughly and examples were given. Both raters were familiar with literature concerning credibility assessment, and had previously participated in a training session concerning CBCA given by a North American expert where example transcripts from a different study were coded. The raters were blind with respect to the hypotheses as well as the condition status of the transcripts.

Personal credibility rating. Before rating the transcripts for CBCA and RM criteria, both raters assessed the overall credibility of the transcribed statements on a 10-point scale based on whether the account was judged to be fabricated or truly experienced (i.e., 1 = freely invented to 10 = self-experienced). This rating was given on a purely intuitive basis where raters were asked not to rely on ratings on the reality criteria.

CBCA. The transcripts were rated for the presence of the first 16 CBCA criteria as delineated by Raskin and Esplin (1991; See Appendix D). Similar to many studies using CBCA in a laboratory setting (e.g., Akehurst, Köhnken, & Hofer, 2001), criteria 17-19 (*self-deprecation, pardoning the perpetrator, and details characteristic of the offence*) were excluded from the analysis as they were not relevant nor applicable to this type of research paradigm. The item *attribution of perpetrator's mental state* was also minimally adapted to become *attribution of other's mental state* with specific reference to the play session

experimenter. Each statement was scored on a 3-point scale (0 = absent, 1 = some indication of presence, 2 = strongly present) for each of the 16 criteria.

Total CBCA Rating. After rating the transcripts for CBCA criteria, both raters assessed the overall credibility of the transcribed statements on a 10-point scale based on the adequacy of the CBCA criteria taking into account the nature and complexity of the play event, and that the children giving the accounts were aged 7 to 8 years. In accordance with Raskin and Esplin's (1991) CBCA instructions, raters were informed that *logical consistency* and *quantity of details* were necessarily present to obtain a high rating.

RM. To assess the presence of Reality Monitoring criteria, the raters coded each account for the presence of eight criteria developed by Sporer (1997) derived from the Judgment of Memory Characteristics Questionnaire (JMCQ) by Sporer and Kuepper (1995)¹. The criteria were *clarity, sensory, spatial, time, emotions, reconstructability, realism* and *cognitive operations*. Each statement was scored on a 3-point scale (0 = absent, 1 = some indication of presence, 2 = strongly present) for each of the criteria. The order of coding was the same for both raters, and was presented as stated above.

¹ The Judgment of Memory Characteristics Questionnaire (JMCQ, Sporer & Kuepper, 1995) was adapted from the Memory Characteristics Questionnaire (MCQ, Johnson, Foley, Suengas, & Raye, 1988) having been modified for the rating of statements by a subject. The subscales in this study were supplemented with brief descriptions of the original JMCQ criteria and the labels for the scale endpoint were revised respectively for individual scales.

RESULTS

Descriptive Aspects and Analyses of Potential Artefacts

To exclude potential artefacts due to characteristics of the accounts (e.g., length), descriptive features of the 60 transcripts were first analyzed. Overall, the accounts varied extensively in length, containing from 7 to 102 sentences ($M = 30.0$, $SD = 17.0$), and consisting of words ranging between 33 and 745 ($M = 296.32$, $SD = 556.47$). Considerable variation was also demonstrated in a measure of average sentence length (i.e., the average number of words per sentence) where accounts ranged from 2.35 to 15.05 ($M = 7.70$, $SD = 2.63$).

Three one-way Analyses of Variance (ANOVA) were conducted to determine whether the number of words, sentences, and average sentence length differed among the three event frequency conditions. Accounts did not significantly differ in number of sentences, number of words, or average words per sentence length (all F s < 1). Thus, any differences found among conditions of event frequency are not a function of variability in amount of available information in the accounts. The means for number of words, sentences, and average word per sentence are shown in Table 2.

Further, three separate MANOVA's were conducted to determine whether there was an effect of age (7, 8), order (1,2,3,4,5), gender (female, male) or interviewer (1, 2) on each of the CBCA and RM criteria. Analyses revealed that there was no effect of age. Despite that multivariate analyses indicated that there was no effect of order [$Wilks' \Lambda = .266$, $multivariate F(4, 21) = .827$, $p = .800$]. Despite null effects at the multivariate level, univariate analyses were conducted as permitted by a priori hypotheses. Univariate analyses revealed an effect of order on the CBCA criterion *mental state of other* [$F(3, 56) = 5.963$, $p < .001$] where a Tukey's post test comparison indicated the criterion was more likely to be rated as present in Order 3 ($M = .40$, $SD = .83$) than the others. There was no multivariate

Table 2.

Mean Number of Words, Sentences, and Average Word per Sentence by Condition

Condition	Words	Sentences	Average Word per Sentence
Single (S-1)	266.95 (168.55)	32.25 (22.77)	8.76 (02.85)
Repeated (S-4)	218.70 (124.55)	28.85 (14.05)	11.67 (18.98)
Fabricated	203.30 (119.38)	28.90 (12.99)	6.84 (02.47)

effect of gender [*Wilks' Lambda* = .521, *multivariate F* (1, 23) = 1.441, *p* = .159]. Univariate analyses indicated there was an effect of gender on the CBCA criterion *quantity of details* [*F* (1, 59) = .6.877, *p* < .05] and RM criterion *clarity* [*F* (1, 59) = 5.033, *p* < .05] where the criteria was more likely to be rated as present for females (*M* = 1.61, *SD* = .097; *M* = 1.27, *SD* = .11, respectively) than males, (*M* = 1.26, *SD* = .086; *M* = 0.89, *SD* = .13). There was also an effect of interviewer on RM criterion *clarity* [*F* (1, 59) = 6.494, *p* < .05] where the criterion was more likely to be elicited by interviewer 1 (*M* = 0.26, *SD* = .64; *M* = 0.82, *SD* = .66) than interviewer 2. Similarly, there was no effect of interviewer at the multivariate level [*Wilks' Lambda* = .585, *multivariate F* (1, 23) = 1.111, *p* = .380]. At the univariate level, *subjective mental state* [*F* (1, 59) = 6.052, *p* < .05] and *cognitive operations* [*F* (1, 59) = 4.872, *p* < .05] were affected by interviewer where the criteria was more likely to be present in accounts elicited by interviewer 2 (*M* = .58, *SD* = .89; *M* = 1.18, *SD* = .96, and *M* = .26, *SD* = .64; *M* = 0.73, *SD* = .98, respectively), than by interviewer 1. *Clarity* was effected by interviewer [*F* (1, 59) = 6.494, *p* < .05] where clarity was more likely to be present in accounts elicited by interview 1 than by interviewer 2. (*M* = 1.26, *SD* = .10; *M* = .82, *SD* = .14). Effects of order, gender and interviewer were controlled by covarying these variables in relevant subsequent analyses.

Incidence of the Ratings

Preliminary descriptive analysis revealed the proportion of reports that contained at least one instance of each of the CBCA and RM criteria varied considerably (see Table 3). The occurrence of CBCA criteria among all the accounts averaged .37 and ranged from .00 present for *unusual details* to .98 present for *quantity of details*. This range of proportions was similar to that reported by Anson, Golding, and Gully (1993). CBCA criterion *unusual details* was never coded to be present in the accounts, and thus was excluded from additional analyses. The endorsement of RM criteria averaged .53 and ranged from .05 present for

Table 3.

*Incidence and Coefficients of Inter-rater Reliability for CBCA and Reality Monitoring
Criteria for Two Raters*

Criteria	Incidence	0,1,2 Coding		Binary Coding		
		<i>r</i>	κ_0	Prop Agree	κ	Max'll RE
CBCA						
1. Logical Consistency	.93	.52*	.32	.90	.44*	.80
2. Unstructured Production	.87	.34	.17	.90	.46*	.80
3. Quantity of Details	.98	.69**	.63**	.95	.95**	.90
4. Contextual Integration	.05	.44*	.29*	.90	.44*	.80
5. Interactions	.71	.54*	.22	.55	.10	.10
6. Reproduction of Conversation	.53	.83**	.68**	.95	.89**	.90
7. Unexpected Complications	.07	.46*	.38*	.85	.32	.70
9. Peripheral Details	.15	.42	.36	.80	.39	.60
10. Misunderstood Details	.03	NA	NA	.95	NA	.90
11. Related Associations	.17	.73**	.68**	.90	.61**	.70
12. Subjective Mental State	.43	.88**	.85**	.95	.90**	.90
13. Other's Mental State	.05	NA	NA	.95	.69**	.90
14. Spontaneous Corrections	.22	.62*	.59**	.90	NA	.80
15. Lack of Memory	.72	.75**	.42**	.65	NA	.30
16. Raising Doubts	.07	NA	NA	.75	NA	.50
Reality Monitoring						
1. Clarity/Vividness	.82	.63*	.52**	.80	.51*	.60
2. Sensory Information	.85	.44	.13*	.95	.69**	.90
3. Spatial Information	.05	.66**	.30*	.90	.55*	.80
4. Time Information	.23	.53*	.41*	.70	.41	.40
5. Emotions	.30	.78**	.66**	.95	.90**	.90
6. Reconstructability	.80	.57**	.47**	.80	.51*	.60
7. Realism	.93	.05	.05	1.00	1.00	1.00
8. Cognitive Operations	.23	.58**	.51**	.80	.61**	.60

Note. NA = Value indeterminate due to the absence of variability in these data.

* $p < .05$; ** $p < .01$ (indicating good agreement).

spatial information to .93 present for *realism*. CBCA and RM criteria varied widely in their rated presence, and thus had varying impacts in diagnosing the accounts (Tully, 1998).

Inter-rater Agreement

Five methods of measuring interrater reliability were calculated and are shown in Table 3. For 0,1,2 coding, Pearson's correlation coefficient (r) and weighted kappa (κ_w) were used. For binary coding (i.e., presence being ratings of 1 or 2 and absence being ratings of 0), proportion of agreement, Cohen's Kappa (κ), and Maxwell's Random Error coefficient (RE) were calculated. Mean agreement between raters for 0,1,2 coding as calculated with the Pearson r was .58 ($SD = .17$), and ranged from .34 for *unstructured production* to .88 for *subjective mental state* for CBCA criteria; and a mean of .53 ($SD = .22$) for RM criteria, ranged from .05 for *realism* to .78 for *emotions*.

To correct for chance, weighted kappas were calculated to evaluate inter-rater reliability. Weighted kappa is a conservative, chance-corrected measure used for ordered categories that takes into account the degree of disagreement between raters (Cohen, 1968). Weighted kappa for CBCA criteria ranged from .17 (slight) for *unstructured production* to .85 (excellent) for *subjective mental state* when linear sets of weights were assigned to the difference levels (Landis & Koch, 1977). For RM criteria, weighted kappa ranged from .05 (poor) for *realism* to .66 (substantial) for *emotions*. Reliability for *unstructured production*, *sensory information* and *realism* were deemed too inadequate, and thus were dropped from remaining inferential analyses involving 0,1,2 coding. An almost complete absence in variability resulted in indeterminate kappa values for *unusual details*, *misunderstood details*, *other's mental state* and *raising doubts* on the CBCA protocol.

Further, the data was dichotomized by recoding all the variables such that ratings of 1 and 2 were combined into a single value representing the "presence" of a criterion, and

ratings of zero were retained to represent the “absence” of a criterion”. Based on these binary data, the average proportion of agreement was .86 ($SD = .12$) for CBCA and ranged from .55 for *description of interactions* to 1.00 for *unusual details*. For RM criteria, mean proportion of agreement was .86 ($SD = .10$) and ranged from .70 for *indications of time* to 1.00 for *realism*. This measure is comparable or greater than agreement reported in previous studies (Anson et al., 1993; Horowitz, Lamb, Boychuk, Krispin, & Reiter-Lavery, 1997; Steller & Köhnken, 1989). Cohen’s Kappa is a chance corrected measure of proportion of agreement, and a special case of weighted kappa where disagreement weights are set equally (Cohen, 1968). Kappa ranged from .00 (slight) for *unusual details* and .95 (near perfect) for *quantity of details* for CBCA criteria. Kappa for RM items ranged from .41 (moderate) for *indications of time* to 1.00 (perfect) for *realism*. Kappa has been criticized for its tendency to become extremely attenuated when the base rate significantly diverges from .50 (Janes, 1979). This limitation is demonstrated in discrepancies between proportion agreement and kappa values, and can be explained by wide variations in incidence.

When the presence of criteria is either extremely rare or extremely common, Maxwell’s Random Error coefficient has been suggested as a more reliable measure of interrater reliability (Anson et al., 1993; Janes, 1979). Maxwell (1977) assumes that if a rater is doubtful about a coding decision, she arrives at the outcome by subjectively “flipping a coin”, rather than making a decision that is ultimately consistent with the base rate, as is assumed when kappa is used. Maxwell’s *RE* coefficient is, thus, a conservative measure of reliability and has values similar to Cohen’s kappa when the incidence is approximately 50 percent. Maxwell’s *RE* coefficient’s for CBCA criteria ranged from .10 (inadequate) for *description of interactions* to 1.00 (perfect) for *unusual details*. *RE* coefficients for RM criteria ranged from .40 for *indications of time* to 1.00 for *realism*. This is considerably higher than that reported in Anson et al. ($RE = .49$) and similar to rates reported by Horowitz

et al. (1997). Based on Maxwell's *RE* values of greater than .50 (adequate reliability), the overall internal consistency was adequate and comparable to other studies assessing these criteria (Anson et al.).

Mean Ratings for CBCA and RM Criteria

The average rating for CBCA and RM criteria were calculated for each of the event frequency conditions and are listed in Table 4. The mean ratings do not vary statistically among the CBCA criteria.

Inclusive Assessments: Personal Credibility and Total CBCA Ratings

Personal Credibility Assessment was converted into a dichotomous variable by ascribing all ratings from 1 to 5 a value as "Not credible" and ratings ranging from 6 to 10 a value of "Credible". This was done to simulate the applied situation where ultimately a judgment must be made about the absolute veridicality of an account. A Chi square analysis for independence was conducted in order to determine whether the levels of event frequency were independent of levels of personal credibility rating. Analyses indicated that levels of event frequency were not independent of personal credibility [$\chi^2 (2, N = 60) = 10.133, p < .01$]. As shown in Figure 1, the SE condition was more likely to be rated as credible in comparison to the RE or FE conditions. Total CBCA Rating was also recoded into a binary variable where ratings from 1 to 5 were coded as "Not credible" and ratings from 6 to 10 were coded as "Credible". This was done because Total CBCA did not discriminate between conditions at the continuous level, and because the determination of credibility by the legal fact finder is either one of credible and not credible. A chi square analysis for independence was not statistically significant [$\chi^2 (2, N = 60) = 1.149, p > .05$].

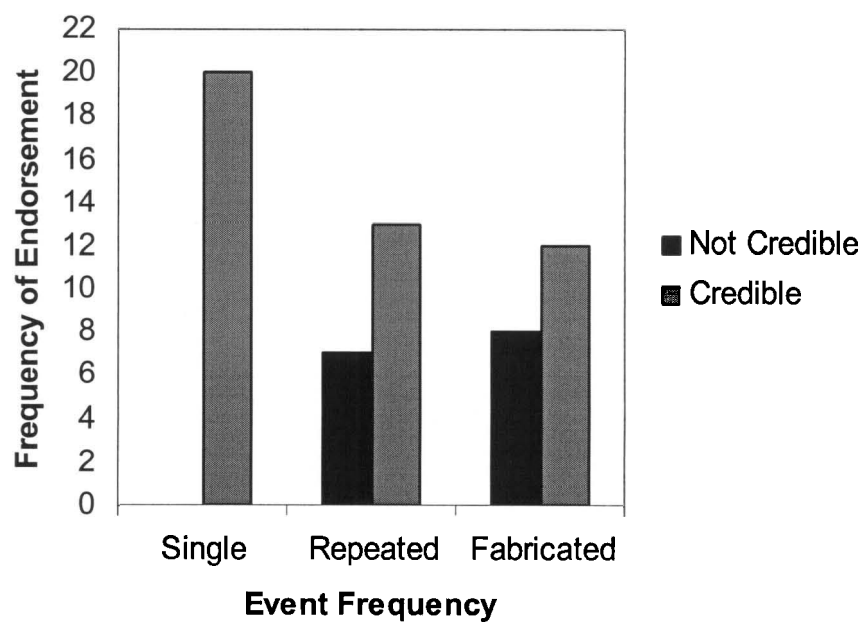
Table 4.

Means of CBCA and RM Criteria as a Function of Event Frequency

Criteria	<u>Event Frequency</u>		
	Single	Repeated	Fabricated
Personal Credibility Rating	8.00	6.30	6.05
CBCA			
1. Logical Consistency	1.75	1.45	1.50
2. Unstructured Production	1.45	1.30	1.30
3. Quantity of Details	1.55	1.45	1.35
4. Contextual Integration	0.01	0.00	0.20
5. Interactions	1.25	1.15	1.00
6. Reproduction of Conversation	0.95	0.90	0.90
7. Unexpected Complications	0.30	0.10	0.00
9. Peripheral Details	0.40	0.25	0.20
10. Misunderstood Details	0.05	0.00	0.05
11. Related Associations	0.40	0.50	.10
12. Subjective Mental State	0.95	0.60	0.85
13. Other's Mental State	0.20	0.00	0.10
14. Spontaneous Corrections	0.60	0.35	0.20
15. Lack of Memory	1.15	1.10	1.35
16. Raising Doubts	0.15	0.15	0.00
Total CBCA Credibility Rating	7.55	6.60	6.30
Reality Monitoring			
1. Clarity/Vividness	1.20	0.95	1.15
2. Sensory Information	1.70	1.65	1.00
3. Spatial Information	0.00	0.00	0.25
4. Time Information	0.45	0.55	0.25
5. Emotions	0.55	0.45	0.60
6. Reconstructability	1.20	0.95	1.05
7. Realism	1.80	1.55	1.35
8. Cognitive Operations	0.50	0.45	0.35

Figure 1.

Perceived Credibility Rating of Single, Repeated and Fabricated Event Accounts



Principal Component Analyses of CBCA Criteria

Principal Component Analysis (PCA) was selected to extract the maximum variance from each component among the CBCA criteria. This extraction method was chosen for its efficacy in reducing a large group of variables into a smaller, more parsimonious number of components for further inferential analyses (Tabachnick & Fidell, 2001). Varimax rotation was used to maximize the variance factor loadings such that high loadings for each component were emphasized and low loadings were minimized (Tabachnick & Fidell, 2001), and ultimately produced a five component solution. The highest eigenvalue was 2.75, the smallest for component 4 was 1.12. Cumulatively, the five components accounted for 63.2% of the variance. Loadings used for component interpretation are in bold and shown in Table 5.

Interpretation of the components based on the factor loadings are as follows:

Component 1, consisting of *unstructured production*, *quantity of details*, and *reproduction of conversation* is generally characterized by structural qualities of the account. Component 2 consists of *contextual integration*, *peripheral details*, and *raising doubts*. This component typically features elements that externally frame the main actions of the storyline. Component 3 is comprised of *unexpected complications*, *related external associations*, and *mental state of other* and are largely characterized by internal processes and attributions. Component 4 consists of *subjective mental state* and *spontaneous corrections* and appears to underpin self monitoring. Component 5 is comprised of *logical consistency*, misunderstood details, and encompasses coherence.

Factor scores developed for the five components were derived using the Bartlett method where factor scores correlate strictly with their own components and are hence unbiased (Tabachnick & Fidell, 2001). In order to determine whether CBCA scores varied as a function of event frequency a one way (Event Frequency: SE, RE, FE) MANOVA,

Table 5.

Factor Loadings of CBCA Criteria

CBCA Criteria	Component 1	Component 2	Component 3	Component 4	Component 5
Logical Consistency	.312	.071	-.034	.214	.777
Unstructured Production	.641	.150	.106	.186	.182
Quantity of Details	.805	.197	.075	.018	.050
Contextual Integration	.078	.725	.042	-.203	.047
Conversation	.834	-.119	.107	-.178	-.143
Complications	-.037	.024	.817	.221	.029
Peripheral Details	.189	.708	.024	.260	-.144
Misunderstood Details	.181	.034	-.096	.258	-.810
Related Associations	.280	.252	.519	-.268	.047
Subjective Mental State	.316	.277	.416	-.478	.176
Mental State of Other	.125	-.249	.596	.032	.171
Spontaneous Corrections	.075	.092	.174	.878	.054
Raising Doubts	.135	.651	-.033	.035	.087

Note. *Unusual details* was excluded due to an absence of variability. *Interactions* and *lack of memory* were excluded due to low reliability.

Table 6.

Mean CBCA Factor Score and Standard Deviation (in parentheses) for Conditions of Event

Frequency

Event Frequency	Component 1	Component 2	Component 3	Component 4	Component 5
Single	0.094 (.952)	.114 (1.33)	.344 (1.45)	.226 (1.07)	-.029 (.990)
Repeated	-.048 (1.05)	-.022 (.647)	-.108 (.699)	-.014 (1.09)	.122 (.603)
Fabricated	-.045 (1.04)	-.092 (.950)	-.236 (.569)	-.212 (.807)	-.093 (1.32)

covarying order and interviewer, was conducted on the derived set of CBCA components. Multivariate analyses indicated that CBCA Components 1, 3, 4 and 5 were not statistically significant [*Wilks' Lambda* = .885, *multivariate F* (2, 3) = .665, $p > .05$]. CBCA Component 2 was marginally significant [F (2, 57) = 2.418, $p = .059$]. Thus, CBCA criteria were not able to discriminate between levels of event frequency (see Table 6 for mean CBCA component scores by condition).

Principal Component Analyses of Reality Monitoring Criteria

A parallel analysis using PCA with Varimax rotation was conducted to extract components among the RM variables. A Scree Plot (i.e., plots eigenvalues against factors) was used to determine that two components existed among the variables, therefore the model was developed for a two component solution. The highest eigenvalue was 2.08, the smallest was 1.33. Overall, the components accounted for 48.6% of the variance. Loadings used to interpret the components are bolded and shown in Table 7.

Component 1 consisted of *clarity*, *emotions* and *reconstructability* and is generally characterized by global form. Component 2 was comprised of *sensory information*, *spatial information*, *realism*, and *cognitive operations*. This component is largely comprised of content cues. A MANOVA was conducted using event frequency (SR, RE, FE) as the independent variable and RM factor scores were dependent variables. Gender and interviewer was covaried in these analyses. Multivariate analyses indicated that there was an effect of event frequency among the components [*Wilks' Lambda* = .760, *multivariate F* (2, 1) = 4.127 $p > .05$]. Analyses indicated that there was no effect of event frequency for Component 1 [F (2, 57) = 1.977, $p > .05$]. An effect of event frequency on Component 2 [F (2, 57) = 4.463 $p < .05$] was revealed, where the SE and RE conditions had a significantly greater presence of Component 2 than the FE condition (see Table 8 and Figure 2).

Table 7.

Factor Loadings of RM Criteria

	Component 1	Component 2
Reality Monitoring	(Internal)	(External)
1. Clarity/Vividness	.826	-.018
2. Sensory Information	.319	.801
3. Spatial Information	.356	-.404
5. Emotions	.519	.058
6. Reconstructability	.862	.057
7. Realism	.186	.687
8. Cognitive Operations	-.143	.373

Note. *Indications of Time* was excluded due to low reliability.

Table 8.

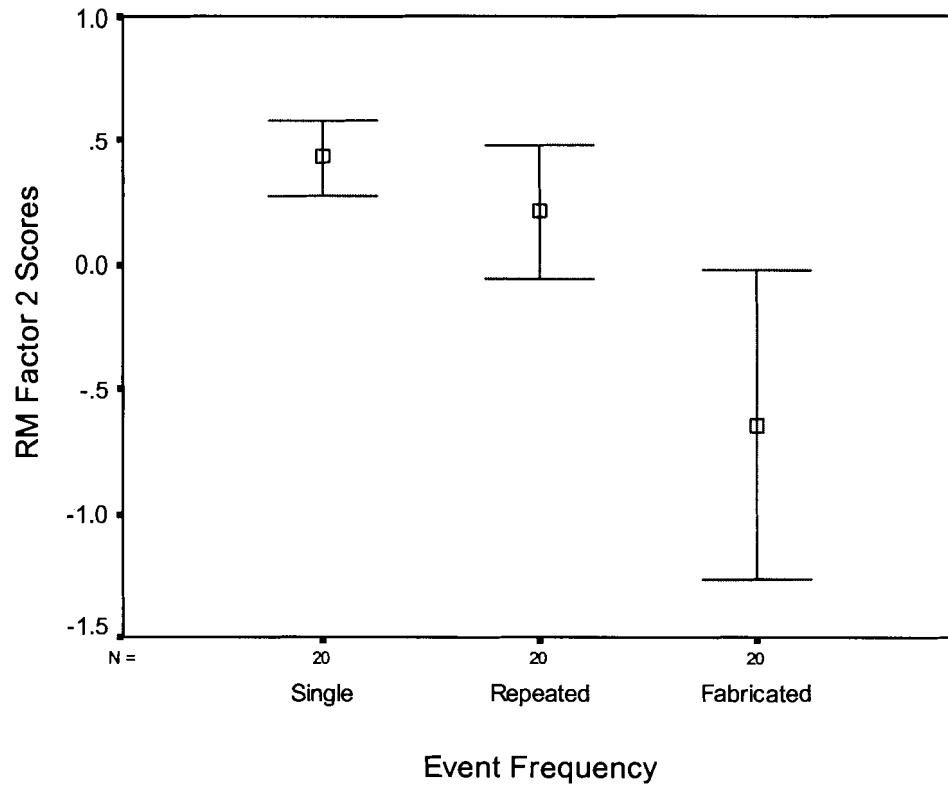
Event Frequency as a Function of Mean RM Component 2 Scores

Event Frequency	Component 1 (External)	Component 2* (Internal)
Single	.142 (.888)	.430 (.341)
Repeated	-.233 (.990)	.213 (.598)
Fabricated	.091 (1.12)	-.643 (1.40)

Note. * $p < .05$

Figure 2.

RM Component 2 (External Cues) Means and Standard Error for Levels of Event Frequency



Note. Bars represent 2 standard errors above and below the mean.

DISCUSSION

The goal of this study was to explore the utility of CBCA and RM as methods in distinguishing between reports of single, repeated and fabricated events. Virtually all of the laboratory work on CBCA has been done comparing memory reports of unique events and reports of fabricated events. Memories for unique events are thought to differ from memories for instances of repeated events, thereby producing differences in surface structure of event reports where repeated events are reported more generally and specific details diminish (but not disappear). This raises the question about the generalizability of protocols that had been developed based on differences in reports of single events to reports of repeated events. However, to the extent that these protocols are used to assess actual reports of child sexual abuse in the field, they are used to discriminate between experienced and fabricated events, regardless of the frequency of experiencing the events. It was unclear whether CBCA and RM criteria would discriminate between authentic and fabricated events irrespective of whether the event was experienced once or repeatedly.

Specific features of an event are more likely to be remembered for unique events than repeated events because they are only experienced a single time (Fivush, 1997). Because specific details are presumed to decrease in reports of repeated compared to reports of unique events, reports of repeated events were postulated to yield statements containing lower CBCA and RM scores on all criteria (except *cognitive operations* as RM theory indicates this should be greater in fabricated accounts) relative to single event reports but higher scores than fabricated reports. There was no reason to expect that reports varying in event frequency would differ on any other criteria except those related to specific details.

In short, the credibility measures were expected to discriminate among the event reports. Due to differences in surface structure based on event frequency where repeated

events are more generalized, it was believed that CBCA and RM protocols would underestimate the authenticity of the account. In the present study, the target events reported by children in the three conditions were similar, thus the discriminative efficacy of both approaches was conducted on comparable material. As discussed below, the credibility measures demonstrated varying degrees of sensitivity in discriminating authenticity and event frequency among the reports.

CBCA Approach

Neither total CBCA Rating, nor derived CBCA components were able to discriminate between fabricated and experienced events, irrespective of the frequency of the experienced event. Similar null effects have been reported by others (Porter & Yuille, 1996; Sporer, 1997; Strömwall et al., in press). There are a number of possible explanations for these findings. First, because the technique was originally designed to discriminate the veridicality of children's reports of sexual abuse, the efficacy of CBCA may be reduced when it is not employed in the field on such claims (Yuille, personal communication). Regardless if this is the case, CBCA is not likely to be largely embraced until it can be shown to reliably discriminate truth status in a controlled environment where noise and confounds can be minimized.

Second, Tully (1998) argues that the reliability and diagnostic ability of CBCA is dependent upon the adequacy (e.g., length, complexity, and prevalence of criteria) of the material being coded. It is possible that the information contained within the transcripts was inadequate to allow reliable discrimination using the CBCA technique (i.e., the incidence of some criteria was below 10%, thereby reducing the full use of CBCA). To test whether adequacy influenced the efficacy of CBCA criteria, the most adequate reports were selected for reanalysis. That is, a median split on number of words was conducted (50th percentile = 207 words), in order to select children producing a high number of words in their recall for

reanalysis. If sufficiency of the transcripts was related to the utility of CBCA criteria, it is reasonable to expect discriminability of the reports that are above the median to be better than discriminability of all reports. A one-way MANOVA using reports that were above the 50th percentile on number of words was conducted on CBCA criteria and the Total Credibility rating where order and interviewer was covaried. While most criteria did not discriminate, analyses indicated that *raising doubts*, [$F(3, 56) = 2.707, p < .05$], *quantity of details* [$F(3, 56) = 4.975, p < .05$], and Total Credibility Rating [$F(3, 56) = 2.938, p < .05$], discriminated between conditions of event frequency. For *raising doubts*, the SE condition had a greater presence than the FE condition (see Table 9). For *quantity of details*, and Total credibility rating, the SE condition evidenced more credibility than the RE and FE condition. Hence, it appears that adequacy is, in fact, related to increased utility of CBCA. These results also indicate that Total CBCA assessment classifies authentic reports based on event frequency differently where accounts of single reports are rated more credibly than those of instances of repeated events which were rated similar to fabrications. These findings imply that the usefulness of the CBCA technique, even for well elaborated accounts, is limited to the extent that it cannot discriminate truthful from coached statements.

Third, Steller, Wellerhaus, and Wolf (1988) suggested CBCA is effective when analyzing an event that involves a) a loss of control, b) the child directly, and c) negative affect. This study involved the child directly, but did not involve a loss of control or negative affect. It is possible that the event, being enjoyable and non-threatening, reduced the discriminative ability of CBCA criteria because it did not involve negative affect. It has been shown, however, that CBCA scores are affected by event familiarity, where CBCA scores are higher for children reporting on a familiar event than events that are unfamiliar (Pezdek et al., 2004). Because children are familiar with playing in general as well as the individual activities in this study, CBCA scores should not have been negatively affected. Notably,

Table 9.

Means and Standard Deviations (in parentheses) of Conditions of Event Frequency as a Function of CBCA Criteria and Total Assessment

Event Frequency	<u>CBCA Criteria / Assessment</u>		
	Raising Doubts	Quantity of Details	Total Credibility Rating
Single ($n = 12$)	.17 (.39)	1.83 (.39)	8.33 (2.02)
Repeated ($n = 10$)	.10 (.32)	1.60 (.52)	6.90 (1.97)
Fabricated ($n = 10$)	.00 (.00)	1.60 (.52)	6.70 (2.75)

similar to the abuse situation (Strömwall et al., in press), participants interacted with a single adult throughout the study.

Fourth, because of the way the fictitious reports were coached, it is possible that the quality of the fabrication was enhanced in relation to uncoached fabrications as reported in the literature. Because a storyboard stimuli was used, it is likely that the children learned to remember the fabricated event verbatim leading to increased memory for specific details over a brief retention period. Given that CBCA is so dependent on reporting specific details, this verbatim recounting of the fabricated event may have boosted the scores on the CBCA measure. In addition, despite the fact the all three groups were told that they would later be asked to describe the event, it is possible that the motivation for the children in the FE group to remember was increased because the participants were being encouraged to “dupe” the interviewer. Further, Ruby and Brigham (1997) suggest that coaching with respect to CBCA items reduces the ability of the rater to correctly classify truth status.

In sum, there are several reasons why CBCA did not discriminate truth status. The stimulus event used may have been too different from the traumatic event report for which the measure was initially designed. Further, the sufficiency of the reports may have been too low to allow adequate utilization of CBCA. Third, the coached fabrications may have been too enhanced for CBCA to detect a difference in authenticity.

RM Approach

The RM approach was more successful in discriminating the veridicality of event reports relative to CBCA. Multivariate analyses of derived RM components indicated that external cues were more likely to be present in accounts based on experience than those that were fabricated. These findings are in line with studies of direct comparisons between CBCA and RM; these studies ultimately found that RM criteria discriminated between true and false reports where CBCA did not (Sporer, 1997; Strömwall et al., in press). *Sensory information,*

spatial information, realism, and cognitive operations were the components that encompassed RM Component 2. These components have been found to have an increased occurrence among reports based on experience than fabrication at a univariate level (Alonso-Quecuty, 1992, 1996; Strömwall et al., in press). The RM approach may have been a better discriminator of authenticity because it is based on an established theoretical perspective as opposed to CBCA which has been criticized for having no such basis (Vrij, 2000). Such a foundation is advantageous because it is informed by the outcomes of numerous empirical studies, which potentially augmented the validity of the protocol.

The RM technique was not sensitive to event frequency. Because script-like memories have been shown, in some cases, to become more pronounced over time (Fivush, 1997), it is possible that the brief retention period used did not allow the memory for the repeated experience memory to schematize. Detecting no difference between single and repeated events can also be explained by Fuzzy trace theory which asserts that the verbatim trace (specific details of the events) fades more quickly than the gist trace (general details) over time (Brainerd & Reyna, 1998). It is possible that the retention time was not long enough for verbatim traces to fade and force the children to rely relatively more on gist memory when providing their account. Further, because the target instance for the RE group was the last in the series, memory for specific details may be explained by the recency effect (i.e., enhanced memory for the last event presented in a serial order; Atkinson & Shiffrin, 1968). Because there was no subsequent event to interfere with memory for the instance, it is possible that details for the last event were remembered better than those instances presented earlier in the sequence (Powell, Thomson, & Ceci, 2003).

In sum, the RM approach discriminated truth status appropriately, and was not sensitive to event frequency. Script theory and Fuzzy Trace theory offer explanations for

these findings given the brief retention period. Remembering the most recent instance is a further explanation for a failure to detect an effect of increased experience.

Perceived Credibility Rating

Interestingly, analyses of perceived credibility assessment indicated the relevancy of event frequency. It was demonstrated in overall perceived credibility assessments, that the SE condition was more likely to be rated as credible in comparison to the RE or FE condition. Clearly, there are implications for the case when genuine experiences (i.e., RE) are found to resemble fabrication. The initial perceived credibility assessment can be likened to a decision made by a juror to the extent that the decision concerning the ultimate truth of a statement that is not based on credibility detection techniques. The juror does not have the benefit of credibility training and is likely to base decisions of credibility on initial instinct (Schuller & Ogloff, 2001). Given that reports of an instance of an authentic repeated experience is initially perceived as resembling a fabricated report, and less credible than a report of a single experience, it is possible that jurors will arrive at the same conclusion. Such a possibility is of interest given that most children who are sexually abused report instances of events that were experienced repeatedly (Connolly & Read, 2003)

Potential Limitations of the Study

The stimuli for coding may have been inadequate to yield a meaningful credibility assessment, however because there are no guidelines concerning sufficiency of statements for CBCA and RM, it is difficult to ascertain if the statements were of adequate length and complexity (Tully, 1998). Second, it is possible that the retention time was not long enough to produce an effect of event frequency. Preliminarily, it appears that studies using retention intervals of one week have had success in obtaining an effect (e.g., Strömwall et al, in press). Third, because the target instance in the second order of the RE condition was not an event that was experienced by the SE or FE conditions, the direct comparison of the three

conditions was restricted. Care should be taken to counterbalance such that the reports reflect the same target instances across conditions for every case to maximize control. A fourth limitation is that the RM criteria was rated after the CBCA criteria and thus a carry-over effect may have occurred where high or low ratings given to CBCA criteria may have influenced RM ratings for related criteria. Analyses indicated, that there were a few significant positive correlations among the criteria (15 out of 91, see Table 10). These outcomes suggest that there was some relationship among the criteria. This finding is tenuous, however, given that ultimately, the RM measure discriminated and CBCA did not suggesting that carry over did not occur. Future studies should strive to counterbalance the rating of the various criteria.

Conclusions and Future Research Directions

Given the large number of child sexual abuse (CSA) cases requiring evidence from young witnesses, coupled with the need to protect clients of the justice system from spurious allegations, the evaluation of credibility and the veracity of children's statements at the investigatory stage of an allegation is a central forensic issue. This study has demonstrated that RM was a better discriminator of authentic reports than CBCA regardless of event frequency. The RM approach has many advantages. Coding is simple to learn in comparison to the extensive training needed for CBCA training (Vrij, 2000). A number of criteria from the CBCA perspective are often discarded in studies due to their rare incidence thereby lowering the discriminative utility of the procedure. On the contrary, RM contains fewer criteria that appear to be more robust, where the criteria is more likely to arise in memory reports (Strömwall et al., in press). Further, RM appears to extend to more general situations, where CBCA was originally formulated for use on typically traumatic events (Strömwall et al., in press). Most importantly, RM rests on an empirically supported foundation, as opposed to CBCA which has been criticized for lacking a solid theoretical foundation, and containing

Table 10.

Correlations of CBCA Criteria and RM Criteria

CBCA Criteria	Reality Monitoring Criteria						
	Clarity	Sensory Info	Spatial Info	Emotions / Feelings	Reconstruct- ability	Realism	Cognitive Operations
Logical Consistency	.237	.262*	.061	.090	.223	.464**	.144
Unstructured Production	.493**	.247	.090	.247	.543**	-.105	.212
Quantity of Details	.666**	.357**	.094	.245	.426**	-.025	.247
Contextual Integration	.079	-.118	.649**	.036	-.022	.061	.067
Reproduction of Conversations	.604**	.216	.104	.272*	.383**	-.097	.185
Unexpected Complications	.059	.112	-.061	.147	.170	.071	.022
Peripheral Details	.145	.176	.118	.121	.096	.112	.064
Misunderstood Details	.110	-.182	-.043	-.117	.118	-.323*	-.100
Related Associations	.199	.188	.103	.352**	.088	-.060	.093
Subjective Mental State	.084	.207	.129	.759**	.150	.085	-.531**
Mental State of Other	.192	.096	-.053	.217	.090	.061	-.124
Spontaneous Corrections	.219	.213	-.117	-.111	.243	.136	-.274*
Raising Doubts	-.040	.112	-.061	-.011	-.124	-.196	.188

Note. * $p < .05$; ** $p < .01$

criteria designed only to verify credibility, and not to detect deception per se (Vrij, 2000). An increasing number of studies demonstrating an inability for CBCA criteria to discriminate authentic and fabricated events in comparison to RM may suggest a review and reformulation of criteria used to identify truth and deception.

Alternatively, it has been suggested that a future integration of the RM and CBCA procedures would be constructive as both approaches offer items that tap non-overlapping dimensions of credibility (Sporer, 1997). To the extent that there are some very strong correlations among the criteria (see Table 10), the removal of items that are contributing to the same variance would yield a more parsimonious measure. Such a union may be useful in the context of further research in the areas discussed below. Certainly, there is a need to test different situations to ensure that the results are generalizable across contexts, and are not entirely tied to one type of event. Also, an examination of the impact of varying retention intervals on the effect of event frequency is warranted. Further exploration of coached fabrication, such as the effect of coaching through different mediums, how coached accounts are represented in memory, and the retention interval of the effect, should be conducted.

Ultimately, it is clear that there is much more work to be done in the area of credibility assessment before such measures are employed in more widespread professional forensic settings. Further investigation of the impact of event frequency on the formulation of memory and subsequent reports of experience is a valid research direction. Overall, results from this study bolster the use of RM both empirically and practically, but do not support the utility of CBCA.

REFERENCES

- Akehurst, L., Köhnken, G., & Hofer, E. (2001). Content credibility of accounts derived from live and video presentations. *Legal and Criminological Psychology, 6*, 65-83.
- Alonso-Quecuty, M. L. (1992). Deception detection and reality monitoring: A new answer to an old question? In F. Lösel D. Bender, & T. Bliesener (Eds.), *Psychology and law: International perspectives* (pp. 328-332). Berlin: Walter de Gruyter.
- Alonso-Quecuty, M. L. (1996). Detecting fact from fallacy in child and adult witness statements. In G. Davies, S. Lloyd-Bostock, M. McMurrin, & C. Wilson (Eds.), *Psychology, law and criminal justice: International development in research and practice* (pp.74-80). Berlin: Walter de Gruyter.
- Anson, D. A., Golding, S. L., & Gully, K. J. (1993). Child sexual abuse allegations: Reliability of criteria-based content analysis. *Law and Human Behavior, 17*, 331-342.
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In W. K. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol 2, pp. 89-195). New York: Academic Press.
- Bala, N. (1999). Child witnesses in the Canadian courts: Recognizing their capacities and needs. *Psychology, Public Policy, and Law, 5*, 323-354.
- Bala, N., Lee, K., Lindsay, R. & Talwar, V. (2000). Legal and psychological critique of the present approach to the assessment of the competence of child witnesses. *Osgoode Hall Law Journal, 38*, 409-451.
- Brainerd, C. L. & Reyna, V. F. (1998). Fuzzy-trace theory and children's false memories. *Journal of Experimental Child Psychology, 71*, 81-129.

- Cashmore, J., & Bussey, K. (1996). Judicial perceptions of child witness competence. *Law and Human Behavior, 20*, 313-334.
- Cohen, J. (1968). Weighed kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin, 70*, 213-220.
- Connolly, D. A., & Lindsay, D. S. (2001). The influence of suggestions on children's reports of a unique experience versus an instance of a repeated experience. *Applied Cognitive Psychology, 15*, 205-223.
- Connolly, D. A., & Read, J. D. (2003). Remembering historical child sexual abuse. *Criminal Law Quarterly, 47*, 438-480.
- Fivush, R., & Slackman, E. (1986). The acquisition and development of scripts. In K. Nelson (Ed.) *Event knowledge: Structure and function in development* (pp. 119-136). Hillsdale, NJ: Erlbaum.
- Fivush, R. (1997). Event memory in early childhood. In N. Cowan (Ed.) *The development of memory in childhood* (pp. 139-161). UK: Psychology Press.
- French, L. A. (1986). The language of events. In K. Nelson (Ed.) *Event knowledge: Structure and function in development* (pp. 119-136). Hillsdale, NJ: Erlbaum.
- Hernandez-Fernaund, E., & Alonso Quecuty, M. L. (1997). The cognitive interview and lie detection: A new magnifying glass for Sherlock Holmes? *Applied Cognitive Psychology, 11*, 55-68.
- Horowitz, S. W., Lamb, P. W., Boychuk, T. D., Krispin, O., & Reiter-Lavery, L. (1997). Reliability of criteria-based content analysis of child witness statements. *Legal and Criminological Psychology, 2*, 11-21.
- Janes, C. L. (1979). Agreement measurement and the judgment process. *Journal of Nervous and Mental Disease, 167*, 343-347.

- Johnson, M. K., Foley, M. A., Suengas, A. G., & Raye, C. L. (1988). Phenomenal characteristics of memories for perceived and imagined autobiographical events. *Journal of Experimental Psychology General*, 117, 371-376.
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, 114, 3-28.
- Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, 88, 67-85.
- Lamb, M. E., Sternberg, K. J., Orbach, Y., Hershkowitz, I., & Esplin, P. W. (1999). Forensic interview of children, In A. Memon & R. Bull (Eds.), *Handbook of the psychology of interviewing* (pp.253-292). Chichester: John Wiley.
- Landis, J. R., & Koch, G.G. (1977) The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.
- Maxwell, A. E. (1977). Coefficients of agreement between observers and their interpretation. *British Journal of Psychiatry*, 130, 79-83.
- Nelson, K. (1986). *Event knowledge: Structure and function in development*. Hillsdale, N.J.: Erlbaum.
- Nelson, K., & Gruendel, J. (1986). Children's scripts. In K. Nelson (Ed.) *Event knowledge: Structure and function in development* (pp. 119-136). Hillsdale, NJ: Erlbaum.
- Orbach, Y., & Lamb, M. E. (2000). Enhancing children's narratives in investigative interviews. *Child Abuse and Neglect*, 24, 1631-1648.
- Pezdek, K., Morrow, A., Blandon-Gitlin, I., Goodman, G. S., Quas, J. A., Saywitz, K. J., Bidrose, S., Pipe, M. E., Rogers, M., & Brodie, L. (2004) Detecting deception in children: Event familiarity affects criterion-based content analysis ratings. *Journal of Applied Psychology*, 89, 119-126.

- Poole, D. A., & Lindsay, D. S. (1995). Interviewing preschoolers: Effects of nonsuggestive techniques, parental coaching, and leading questions on reports of non-experienced events. *Journal of Experimental Child Psychology, 50*, 129-154.
- Porter, S., & Yuille, J. C. (1996). The language of deceit: An investigation of the verbal clues to deception in the interrogation context. *Law and Human Behavior, 20*, 443-458.
- Porter, S., Yuille, J. C., & Lehman, D. R. (1999). The nature of real, implanted, and fabricated memories for emotional childhood events: Implications for the recovered memory debate. *Law and Human Behavior, 23*, 517-537.
- Powell, M. B., Roberts, K. P., Ceci, S. J., & Hembrooke, H. (1999). The effects of repeated experience on children's suggestibility. *Developmental Psychology, 35*, 1462-1477.
- Powell, M. B., & Thomson, D. M. (1997). Contrasting memory for temporal-source and memory for content in children's discrimination of repeated events. *Applied Cognitive Psychology, 11*, 339-360.
- Powell, M. B., & Thomson, D. M. (1996). Children's memory of an occurrence of a repeated event: Effects of age, repetition, and retention interval across three question types. *Child Development, 67*, 1988-2004.
- Powell, M. B., Thomson, D. M., & Ceci, S. J. (2003). Children's memory of recurring events: Is the first event always the best remembered? *Applied Cognitive Psychology, 17*, 127-146.
- Price, H. L., & Connolly, D. A. (in press). Event frequency and children's suggestibility: A study of cued recall responses. *Applied Cognitive Psychology*.
- Raskin, D. C., & Esplin, P. W. (1991). Statement validity assessment: Interview procedures and content analysis of children's statements of sexual abuse. *Behavioral Assessment, 13*, 265-291.

- 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.
- Santtila, P., Roppola, H., Runtti, M., & Niemi, P. (2000). Assessment of child witness statements using criteria-based content analysis (CBCA): The effects of age, verbal ability, and interviewer's emotional style. *Psychology, Crime and Law*, 6, 159-179.
- Saywitz, K., & Camparo, L. (1998). Interviewing child witnesses: A developmental perspective. *Child Abuse and Neglect*, 22, 825-843.
- Schuller, R. A., & Ogloff, R. P. (2001). *Introduction to psychology and law: Canadian perspectives*. Toronto: University of Toronto Press.
- Shank, R. & Abelson, A. (1997). *Scripts plans goals and understanding*. Hillsdale, NJ: Erlbaum.
- Sporer, S. L. (1997). The less traveled 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. & Kuepper, B. (1995) Realitätsüberwachung und die Beurteilung des Wahrheitsgehaltes von Erzählungen: Eine experimentelle Studie (Reality monitoring and the judgment of credibility of stories: An experimental investigation.) *Zeitschrift für Sozialpsychologie*, 26, 173-193.
- Steller M., & Köhnken, G. (1989). Criteria-Based Statement Analysis. In D. C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence* (pp. 217-245). New York: Springer Publishing Company.
- Steller, M., & Wellershaus, P. (1996). Information enhancement and credibility assessment of child statements: The impact of the cognitive interview technique on criteria-based content analysis. In G. Davies, S. Lloyd-Bostock, M. McMurrin, & C. Wilson (Eds.),

Psychology, law and criminal justice: International development in research and practice (pp.74-80). Berlin: Walter de Gruyter.

- Strömwall, L. A., Bengtsson, L., Leander, L., & Granhag, P. A. (in press). Assessing children's statements: The impact of a repeated experience on CBCA and RM Ratings. *Applied Cognitive Psychology*.
- Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics*, 4th Edition. Needham Heights, MA: Allyn & Bacon.
- Tully, B. (1998). Reliability of criteria based content analysis of child witness statements: Cohen's kappa doesn't matter. *Legal and Criminological Psychology*, 3, 183-188.
- Undeutsch, U. (1989). The development of statement reality analysis. In J. C. Yuille (Ed.), *Credibility assessment* (pp.101-120). Dordrecht, The Netherlands: Kluwer.
- Vrij, A. (2000). *Detecting lies and deceit: The psychology of lying and the implications for professional practice*. Chichester: John Wiley.
- Vrij, A. (in press). Criteria-based content analysis: A qualitative review of the first 37 studies. *Psychology, Public Policy and Law*.
- Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2002). Will the truth come out? The effect of deception, age, status, coaching, and social skills on CBCA Scores. *Law and Human Behavior*, 26, 261-283.
- Vrij, A., Edward, K., Roberts, K. P., & Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal Behavior*, 24, 239-263.
- Yuille, J. C. (1988). The systematic assessment of children's testimony. *Canadian Psychology*, 29, 247-262.

APPENDIX A: LETTER OF ETHICAL APPROVAL

SIMON FRASER UNIVERSITY

OFFICE OF RESEARCH ETHICS



BURNABY, BRITISH COLUMBIA
 CANADA V5A 1S6
 Telephone: 604-291-4370
 FAX: 604-291-4860

July 5, 2002

Ms. Jennifer Lavoie
 Graduate Student
 Department of Psychology
 Simon Fraser University

Dear Ms. Lavoie:

**Re: The Language of Memory: The Differences in Content Criteria of
 Children's Verbal Reports of Fabricated Unique and
 Instances of Repeated Experience**
*"Linguistics Markers of Reported Fabricated, Single and
 Repeated Events in Child Witnesses"*
 NSERC

I am pleased to inform you that the above referenced Request for Ethical Approval of Research has been approved on behalf of the Research Ethics Board. This approval is in effect for twenty-four months from the above date. Any changes in the procedures affecting interaction with human subjects should be reported to the Research Ethics Board. Significant changes will require the submission of a revised Request for Ethical Approval of Research. This approval is in effect only while you are a registered SFU student.

Best wishes for success in this research.

Sincerely,

Dr. Hal Wejnberg, Director
 Office of Research Ethics

c: D. Connolly, Supervisor

/bjr

SIMON FRASER UNIVERSITY

OFFICE OF RESEARCH ETHICS
ROOM 2105 STRAND HALL



BURNABY, BRITISH COLUMBIA
CANADA V5A 1S6
Telephone: 604-291-3447
FAX: 604-268-6785

July 31, 2003

Ms. Jennifer Lavoie
Graduate Student
Department of Psychology
Simon Fraser University

Dear Ms. Lavoie:

**Re: The Language of Memory: The Differences in Content Criteria of
Children's Verbal Reports of Fabricated Unique and
Instances of Repeated Experience**

*"Linguistics Markers of Reported Fabricated, Single and
Repeated Events in Child Witnesses"*

NSERC
Extension

In response to your request dated July 30, 2003, I am pleased to approve, on behalf of the Research Ethics Board, the extension to July 5, 2004, of the above referenced Request for Ethical Approval of Research originally approved on July 5, 2002.

Best wishes for continued success in this research.

Sincerely,

Dr. Hal Weinberg, Director
Office of Research Ethics

c: Dr. D. Connolly, Supervisor

/jmy

APPENDIX B: INSTRUCTIONS FOR FABRICATION

Orienting Comment

Hi. My name is _____. Can you say your name? How are you?

How do you like school?

What are your favorite things to do?

Instructions

Today we are looking at how children tell stories and I was hoping that you might help me find out more about story telling. I'm going to ask you to make up a story using some pictures that I have here and this tape recorder will pick up what you say, OK?

We are trying to see how good your story telling skills are and how well you tell a believable story. These are pictures of a day called APRON DAY. I would like you to make up a story about you and me and APRON DAY using these pictures.

Tomorrow a girl named _____ will be coming in to ask you about APRON DAY. The game is to try your best to convince _____ to believe that the story you make today actually happened in real life.

Do you understand? Are you ready to start?

Story Generation

OK. Let's make up a story about you and me on APRON DAY.

- Show pictures

If you want, close your eyes and try to imagine that you are in the pictures doing all of these things. Think about what would be happening and how you think things would be.

- Clarify misinterpretation of critical details only
- Have child say whole story again using the pictures

Conclusion

That was great, thank you very much for making a story. You tell stories really well. Remember that tomorrow _____ will be coming in to ask you about APRON DAY. When you are telling your story, try not to say anything about these pictures because we are trying to make this a believable story. To help us learn more about story telling, remember to try your best to make her believe that your story really happened. Do you have any questions?

APPENDIX C: MEMORY INTERVIEW PROTOCOL

Subject # _____

Date: _____

Memory Interview

Orienting Comment

1. Hello, my name is _____. I'm interested in learning more about APRON DAY today. I'm going to ask you some questions, and I might write down some notes so that I can remember what you say. This tape recorder will pick up what you say too. Is it OK if we record what we say? Remember to talk nice and loud, OK? If you need to take a break then just let me know, OK?

Rapport Questions

2. How do you like day camp/school?
3. What are your favourite things to do?

Event Introduction and Free Recall

4. I'm going to ask you now about what happened during APRON DAY, the day that Jen wore the apron. I'd like you to tell me everything you can remember.

a) What happened on APRON DAY?

- Don't interrupt. When the child stops talking give her/ him about 15 seconds before prompting.
- You are free to use **non-leading** prompts to **clarify** information

b) What else happened on APRON DAY? If you want, you can close your eyes and think about it. I can give you lots of time to think about it.

c) Do you remember anything else? Even things you don't think are important?

d) Is there anything else that you can remember that you haven't told me?

Specific Cued-Recall Question Introduction

Now I have some more specific questions for you. I want you to think about APRON DAY. For the following questions, it's OK to say "I don't know" if you don't remember. If I ask you something that you told me about earlier, it doesn't mean that it's wrong. It is just that I have to ask you all of these questions and write down what you say. OK?

8. You did a warm-up exercise during APRON DAY, can you tell me about the exercise you did?

9. You put together a puzzle during APRON DAY, what picture was on the puzzle?

10. During APRON DAY you listened to music when you made the puzzle, what instrument did you listen to?

11. What kind of hat did you wear during APRON DAY?

12. You learned a magic trick during APRON DAY, what trick did you learn?

13. What magic words did you say to help you with the magic?

14. What did you use to help you with the magic?

15. You got a sticker during the science game, what picture was on the sticker?

16. When you played the science game during APRON DAY, what did you learn about?

17. What did your materials that you used for the science game smell like?

Interview Conclusion

18. Thanks for helping me understand more about what happened during APRON DAY.
You did a great job. You remembered lots of things. Do you have any questions for me?

APPENDIX D: CBCA AND RM CODING PROTOCOL

Part A: Personal Assessment

Personal Credibility Rating: Judge the credibility of the account on a 10 point scale, whether you believe the account was invented or self-experienced. This rating is to be given on a purely intuitive basis, without paying particular attention to the ratings of the individual criteria. Altogether, was the event according to your own personal estimate freely invented or personally experienced (considered independent of the evaluations)?

Part B: Statement Validity Assessment Criteria-Based Content Analysis (CBCA) Adapted from D.C. Raskin, P. W. Esplin (1991)

The criteria refer to the core action of a story (that is, it does not refer to earlier or later mentioned events). All criteria are rated on a scale of:

- 0 = does not exist
- 1 = exists somewhat- certain sign exist
- 2 = definitively exists

GENERAL CHARACTERISTICS

Assessing the presence or absence of the criteria under the General Characteristics category requires an examination of the statement as a whole. Logical consistency and quality of details are expected in all valid statements.

- *1. **LOGICAL CONSISTENCY:** Is the statement coherent? Is the content fundamentally logical? Do the different segments fit together? (Note: Peculiar or unique details or unexpected complications do not diminish logical structure).
2. **UNSTRUCTURED PRODUCTION:** Are the descriptions unconstrained? Is the report somewhat unorganized? Are these digressions or spontaneous shifts of focus? Are some elements distributed throughout? Not overly structured/chronological (Note: This criterion requires the account is logically consistent. This criterion is important but not an absolute requirement).
- *3. **QUANTITY OF DETAILS:** General impression of the number of shown details of the core action. Is the event meaningfully described and illustrated? Are these specific descriptions of place or time? Are persons, objects, and events specifically

described? The child should give enough detailed information to provide a basis for knowing the specific nature of the acts and some of the surrounding circumstances (Note: Repetitions are not counted.)

SPECIFIC CONTENTS

Specific Contents refers to the particular passages in the statement. Such contents provide the concreteness and vividness characteristic of actually experienced events. The first 2 specific criteria (Contextual Integration and Description of Interactions) are especially important.

***4. CONTEXTUAL INTEGRATION:** Are events placed in spatial and temporal context? Is the action connected to other incidental events, such as routine daily occurrences? The presence of this criterion requires more than a minimal description of context of the type, "It happened at school with my friends".

***5. DESCRIPTIONS OF INTERACTIONS:** Are there reports of actions and reactions or conversation composed of a minimum of three elements? (Note: Verbatim reproduction of conversation is also scored under criterion). E.g., "She gave me a marker [action by experimenter], so I began to draw on the paper [reaction by witness], and she told me I was doing a Great Job! [action by experimenter]".

N.B. Other specific contents are expected to occur relatively less frequently, even in valid accounts. Thus, the validity of a statement is strongly supported by the presence of some criteria, but their absence does not necessarily invalidate the statement. This is especially true for many of the remaining criteria.

6. REPRODUCTION OF CONVERSATION: Is speech or conversation during the event reported in its original form? This criterion may be absent in a substantial proportion of statements, but its presence can be very compelling. (Note: Use of unfamiliar terms or quotes are especially strong indicators, even when attributed to only one participant. A child who has not experienced the event is unlikely to provide what appears to be a verbatim reproduction of an utterance that is expressed in terms of the different speakers, such as incorporating adult expressions).

7. UNEXPECTED COMPLICATIONS: Was there an unplanned/unforeseen interruption or an unexpected complication or difficulty during the event? This criterion refers to contents that a child is unlikely to invent. E.g., "I hit my head on the table".

8. UNUSUAL DETAILS: Are there details of persons, objects, or events that are unusual, yet meaningful in this context? (Note: Unusual details must be realistic.) This criterion refers to the details that are odd, yet meaningful, for the context; for the context; or don't appear often in accounts. [E.g., references to allergies to scented

markers; having a broken arm and having difficulty in participating in event activities].

9. **PERIPHERAL DETAILS:** Are peripheral details described in connection with the event/situation that are not essential, and do not contribute directly to the specific core action? (The course of action would be understood without these details). [E.g., Her coat was wet when she came into the classroom because it was raining outside; my friends were playing outside the window].
10. **ACCURATELY REPORTED DETAILS MISUNDERSTOOD:** Did the child correctly describe an object or event but interpret it incorrectly? This criterion is infrequently encountered.
11. **RELATED EXTERNAL ASSOCIATIONS:** Is there reference to an event/conversation that is related in some way to the incident, but is not part of the incident? This criterion involves references to events that are not within the boundaries of the present incident but shares features with it. [E.g., references to the experimenter discussing conducting the same event with other children in the class; the child mentioning that he has outside experience doing one of the activities].
12. **SUBJECTIVE MENTAL STATE:** Did the child describe feelings or thoughts experienced at the time of the incident (Note: This criterion is not satisfied when the witness responds to a direct question, unless the answer goes beyond the question)? [E.g., "I thought the game was fun].
13. **ATtribution OF OTHER'S MENTAL STATE:** Is there reference to the experiment's feelings or thoughts during the incident (Note: Descriptions of overt behavior do not qualify). [E.g., "She was happy with my picture].

MOTIVATION-RELATED CONTENT

Motivation-Related Contents refers to motives of the witness in the sense that a fabricating witness is not expected to incorporate such contents into the account. Thus, a lying or coached witness is expected to attempt to maintain the basic story without modification, to try to answer all questions even if that requires additional fabrication, and not raise doubts about the believability of the story, generally.

14. **SPONTANEOUS CORRECTIONS:** Were corrections offered or information added to material previously provided in the statement? Were additions or corrections spontaneously offered? (Note: Responses to direct questions do not qualify). [E.g., "There were 3 rocks, no, there were actually 4 rocks 'cause I put them into two piles.]
15. **ADMITTING LACK OF MEMORY:** Did the child indicate lack of memory or knowledge of an aspect of the incident? (Note: In response to a direct question, the answer must go beyond "I don't know" or "I can't remember.") [E.g., "The house was

blue - I forget the rest"; "I don't remember anything about this except the game was about a fish].

16. RAISING DOUBTS ABOUT ONE'S OWN TESTIMONY: Did the child express concern that some part of the statement seems incorrect or unbelievable? (Note: Merely asserting that one is telling the truth does not qualify.) Because raising doubt implies a possibility of insincerity, a child who is fabricating will generally avoid doubt, whereas a child who is reporting truthfully may recognize the implausibility of some aspects of the event they are describing. [E.g., You know, it sounds weird, but she was feather on her head.]

Part C: Total CBCA Assessment

CBCA Credibility Rating: Judge the credibility of the account on a 10 point scale, whether you believe the account was invented or self-experienced based on the CBCA criteria. [Logical consistency and quantity of details must be present, taking into account the nature and complexity of the play event, and that the children giving the accounts are aged 7-8.]

Part D: Reality Monitoring Criteria
Judgment of Memory Characteristics Questionnaire (JMCQ)
Subscales: Sporer and Kuepper (1995)

INSTRUCTIONS: The criteria refer to the core action of a story (that is, it does not refer to earlier or later mentioned events). All criteria are rated on a scale of:

0 = does not exist

1 = exists somewhat- certain sign exist

2 = definitively exists

(1) CLARITY/VIVIDNESS

This criterion is concerned with clarity and vividness of the memory and/or its representation. Possible information for this criterion to achieve a rating of high would include details regarding amount of visual details, accuracy, comprehensibility (i.e. of the sequence) and vividness of the representation

Scale: Low - high

- a) Clarity: In your opinion, the event remains how clear in the memory of the person telling the story?
- b) Quantity of visual details:
How many visual details are described?
- c) Vividness: How vividly is the event described?
- d) Precision of details: Are details described only superficially or very precisely?
- e) Order of events: How comprehensible is the order of events described?

(2) SENSORY INFORMATION

This criterion is concerned with the information regarding the senses and the number of such indications. Possible information for this criterion to achieve a rating of many would include details about colours, noises and sounds (apart from speaking), smells, tastes, touch (i.e., how objects feel).

Scale: None - many

- a) Colours: Are objects, persons or the environment described without color or are colors described?
- b) Sounds: How many or how intensely are sounds and tones mentioned?
- c) Smells: How many or how intensely are smells mentioned?
- d) Touch: How often or how intensely is something described when it is being touched?
- e) Taste: How often or how intensely are tastes described?

(3) SPATIAL INFORMATION

This criteria is concerned with indications of spatial information and precision of describing spatial conditions. Possible details to include for this criterion to be rated as precise could be about the environment, the specific location/place of core action, and/or the spatial arrangement of persons and objects.

Scale: Not precise - precise

- a) Location: How clearly is the location of the event described?
- b) Setting: How familiar does the environment and the setting of the event appear to be to the person?
- c) Spatial arrangement of objects: How clearly is the spatial arrangement of objects described?
- d) Spatial arrangement of persons: How clearly is the spatial arrangement of persons described?

(4) INDICATIONS OF TIME

This criterion is concerned with general indications of the time. Possible information to include for this criterion to achieve a rating of precise would be details concerning the year, season, day or hour.

Scale: Not precise - precise

- a) Time: How clearly is the time of the event described?
- b) Year: How clearly is the year of the event described?
- c) Season: How clearly is the season of the event described?
- d) Day: How clearly is the day of the event described?
- e) Hour: How clearly is the hour of the event described?
- f) Age of the storyteller: How old was the person at the time of the event?
- g) Duration: How long did the event last?

(5) EMOTIONS AND FEELINGS

This criterion is concerned with the description of emotion. Possible information to include for this criterion to achieve a rating of many would include details about feelings, intensity of feelings during the event, feelings at the time of telling, played role of the storyteller and the possibility of being able to draw conclusions about the personality of the storyteller.

Scale: None - many

- a) Evaluation of event (tone of the event): Which overall tone did the event have for the person? Did the person experience it rather negatively or rather positively?
- b) Role-played in the event: Which role did the person play in this event: More like a spectator or more like a participant?
- c) Remembered feelings: How well does the person remember feelings at the time of the event?
- d) Type of feelings: At the time of the event, were the feelings negative or positive?

- e) Intensity of feelings at the time: At the time of the event, were the feelings weak or intense?
- f) Intensity of feelings now: At the time of telling the story, were the feelings weak or intense?
- g) Implications regarding personality: How much does the story reveal about the personality of the person?

(6) STORY RECONSTRUCTABILITY

This criterion is concerned with the degree to which the action can be reconstructed. Possible information to include for this criterion to achieve a rating of precise would include details concerning complexity of the action, actual and assumed consequences, and no doubt about their own memory.

Scale: Not precise – precise

- a) Complexity of story line: How simple or how complex is the story line?
- b) Presumed consequences: At the time the event occurred, did it seem to have serious implications?
- c) Factual consequences: According to the account, did the event have serious consequences or implications that became clear afterwards?
- d) Quality of remembering: How well does the person seem to remember the event?
- e) Description of previous events: Is the event embedded into a broader context by describing events which took place before the event described?
- f) Description of subsequent events: Is the event embedded into a broader context by describing events which took place after the event described?
- g) Doubts about the accuracy of remembering: Does the storyteller have any doubts about the accuracy of the memory for the event or is he/she rather sure?

(7) REALISM

This criterion is concerned with the degree to which the narration is realistic. Possible information to include for this criterion to achieve a rating of realistic would include details concerning likeliness of the event, extraordinariness, incredible details and believability.

Scale: Not realistic – realistic

- a) Realism of story line: How realistic is the story line? Does it appear bizarre or realistic?
- b) Likelihood of event: Could an event like this have happened to you in a comparable way?
- c) Extraordinariness of event: To what extent is the story surprising, unpredictable or extraordinary?
- d) Incredible details: To what extent does the story contain incredible details?
- e) Believability: If someone else told you the story as the storyteller did, how likely would you believe it?

(8) COGNITIVE OPERATIONS

This criterion is concerned with descriptions of cognitive operations. Possible information to include for this criterion to achieve a rating of many would include details concerning thoughts during the event, thoughts on preceding or following events, repeated consideration and telling of the event.

Scale: None -many

- a) Thoughts: How precisely are thoughts described which the person had at the time of the event?
- b) Repeated thinking about the event: How often has the storyteller thought about the event?
- c) Repeated talking about the event: How often has the storyteller talked about the event?

Criteria were translated from Sporer, S. L. and Kuepper, B. (1995) *Realitätsüberwachung und die Beurteilung des Wahrheitsgehaltes von Erzählungen: Eine experimentelle Studie* [Reality monitoring and the judgment of credibility of stories: an experimental investigation.] *Zeitschrift für Sozialpsychologie*, 26, 173-193. This questionnaire is a modified version of the Memory Characteristics Questionnaire (MCQ) by Johnson, Foley, Suengas and Raye (1988).