CRITERIA-BASED CONTENT ANALYSIS:
AN EXPERIMENTAL INVESTIGATION WITH CHILDREN

by

RISHA D. JOFFE

B.Sc.(hon.), The University of Calgary, 1981

M.A., The University of British Columbia, 1986

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Psychology)

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

December, 1992

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

(Signature) _____

Department of Psychology

The University of British Columbia
Vancouver, Canada

Date Dec. 29/92

Abstract

The aim of the present study was to experimentally test the Undeutsch Hypothesis, which holds that children's statements based on self-experienced events are qualitatively and quantitatively different from statements based on coaching. Specifically, this study tested the validity of Criteria-Based Content Analysis (CBCA, a system for assessing the credibility of eyewitness reports) for discriminating between credible and noncredible eyewitness reports by children. As well, two other tests of the quantitative and qualitative differences between credible and noncredible eyewitness reports were included. One hundred and forty-two children (74 Grade 4, 68 Grade 2) were tested in three conditions: (1) Live Event, in which children were actively involved in a staged event (the event was complex and included many features considered relevant to credibility), (2) Heavily Coached, in which children did not experience the event but were told in detail about it (including details which, if reported, would be assigned significance by CBCA), and (3) Lightly Coached, in which children did not experience the event but were provided with a brief account of it, with the expectation that they would fill in details to make their reports believable. Children were asked to recall the event in individual interviews. Transcribed interviews were evaluated using CBCA. Results of the study provided mixed support for the Undeutsch Hypothesis. For Grade 4 children, CBCA significantly discriminated between the Live Event and Lightly Coached conditions, but not between the Live Event and Heavily Coached conditions. Thus,

although CBCA accurately distinguished credible from lightly coached reports by this older group of children, reports of the heavily coached children fooled CBCA evaluation. For Grade 2 children, CBCA did not discriminate between the three conditions. This result raised questions about the applicability of CBCA to the reports of younger children. Results of the other two tests of quantitative and qualitative characteristics indicated that these systems did not aid in discriminating between credible and noncredible reports. The implications of these findings for the empirical validation of CBCA and for the use of this system in making credibility decisions in the forensic context are discussed. At this point in time, the assessment of CBCA is still taking place. Until further testing is completed, CBCA should be viewed as one approach to credibility assessment that has clinical support but limited empirical validation.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I warmly thank my research supervisor, John Yuille, for his steady guidance and support throughout the 3-1/2 years I worked on this project. I am grateful to you, John, for your role in allowing me to bring this thesis to completion, and for helping me to see the 'bigger picture' by providing me with the opportunity to present the findings of this study at the NATO-ASI in Italy. Thanks John, I feel lucky to have been your student.

I greatly appreciate the helpful advice and input of the other two members of my thesis advisory committee, Charlotte Johnston and Peter Graf. I enjoyed our contact, and both of you played an important role in expanding my thinking on the issues involved in this research.

I heartily thank Ralph Hakstian for statistical consulting (in formally scheduled meetings and as a helpful office neighbor). Thanks also to Marsha Schroeder for sharing her expertise on Generalizability Analysis, and to Linda Scratchley for addition statistical consulting (Spitballs, etc.).

Thanks very much to Robin Hunter and JoAnn Miller for serving as the experimenter, and to David Marxsen--repairman extraordinaire--for being my reliable confederate actor. Thank you to Patricia Tollestrup, Sonja Pietsche, Esther Chetner, Kiersten Humphrey, Leta Labiuk, and Judy Zaparniuk for conducting interviews with the children. As well, I am grateful to JoAnn Miller and Dana Sair for transcribing the interviews.

I express my deep gratitude to the volunteers who gave willingly of their time to learn the coding procedures, attend weekly coding meetings, and on their own time apply the coding procedures to the transcribed interviews. So, a heart-felt thanks to the following groups of people: Nadine Dodd, Ila Doyle, Christine Saunders, and Sabrina Sealaus for CBCA evaluation; Kim Behrenz, Brigitta McMillan, and Josephine Tse for Amount and Accuracy of Detail coding; and Heather Dick, Marc Keith, and Kerry Lau for coding of the Johnson/Schooler qualitative characteristics.

I would like to acknowledge John Turtle for helpful discussions in the planning stage of this project, and in particular for bringing to my attention the relevance of work conducted by Johnson and by Schooler.

I am grateful to the Vancouver and Richmond School Boards, and in particular to Grade 4 and 2 teachers and students at the following elementary schools: False Creek, Franklin Community, Jamieson, Kerrisdale, Kerrisdale Annex, Laurier Annex, MacCorkindale in Vancouver; and Dixon and Diefenbaker in Richmond.

Finally, I thank my family and friends for being there and encouraging me as I worked my way through this project.

Chapter 1

OVERVIEW OF THESIS

Children's eyewitness testimony has historically been mistrusted by the judicial system, but research carried out in the past 20 years has demonstrated convincingly that children are capable of providing credible testimony. Nevertheless, on occasion, children do present false testimony. There is a need for a system to distinguish between credible eyewitness reports (i.e., those based on self-experienced events) and non-credible eyewitness reports (i.e., those based on fantasy/coaching).

The Undeutsch Hypothesis states that eyewitness reports based on true experience are qualitatively and quantitatively different from statements based on fantasy/coaching. Undeutsch proposed a number of qualitative characteristics differentiating real and coached/invented memories. On the basis of these proposed qualitative differences, Statement Validity Analysis (SVA)--comprised of a Criteria-Based Content Analysis (CBCA) procedure and a Validity Checklist--was developed as a procedure for assessing the credibility of child witness's statements. SVA is increasingly being used by psychologist experts, police officers, and social workers to aid in assessments of the credibility of child witness's reports. In spite of the growing popularity of this procedure, its reliability and validity as a credibility assessment tool have not been adequately tested.

In order to adequately test this procedure, a combination of field and experimental research is required. Each type of research

has advantages and disadvantages. Cases assessed in field investigations are better suited to SVA's qualitative evaluation, because the procedure was designed with reference to the type of emotional/personal dynamics associated with eyewitness events in the forensic context. In field investigations, credibility decisions based on SVA are compared with the conclusions drawn from other evidence in the case (e.g., in sexual abuse cases, disclosure by the alleged abuser, physical or medical evidence, etc.). Although such field research is critical, it alone is insufficient for drawing conclusions about the validity of SVA-based credibility decisions. This is because in such forensic cases, corroborating evidence (e.g., confession by the alleged abuser, physical/medical evidence) is often lacking. Thus, there is often no way of knowing the objective facts regarding what did or did not take place.

Experimental investigations, on the other hand, make it possible to know with certainty whether a given statement is based on personal experience simply by virtue of the assignment of children to experimental conditions (e.g., live event or coaching), and the controlled nature of the event or coaching to which children in the various conditions were exposed. This absolute knowledge of children's experimental condition allows judgments made using statement analysis to be compared to the objective criterion of group membership. However, this type of research alone is also insufficient as a test of SVA. For ethical reasons, experimenters cannot, and should not, duplicate the degree of emotional involvement or trauma experienced by children in real-life events leading to eyewitness reports. Nor should they attempt to recreate

the web of motivational factors operating on child witnesses at the time of giving testimony in the forensic context. Further, not all components of SVA (e.g., a number of criteria relating specifically to features of sexual abuse scenarios, and the Validity Checklist) can be applied to children's statements obtained in experimental investigations. Thus, both field and experimental research are needed. The findings from both must be integrated in order to reach conclusions about the validity of SVA.

The present study is an experimental investigation designed, in general, to test the Undeutsch Hypothesis, which holds that children's statements based on self-experienced events can be distinguished from statements based on coaching on the basis of qualitative and quantitative characteristics. Specifically, the methodology employed allowed a rigorous test of the reliability and validity of the CBCA component of SVA for making decisions about the credibility of children's statements. CBCA-based credibility decisions were obtained for transcribed eyewitness reports by children in three experimental conditions:

1. Live Event (LE), in which children were actively involved in a staged event. The event was complex and included many features considered relevant to credibility.

2. Heavily Coached (HC), in which children did not experience the event, but were told in detail about it. This coaching was designed to incorporate details of the event which, if reported, would be assigned credibility enhancing significance by CBCA. Children were asked to pretend that they had actually experienced the event and to attempt to fool the interviewer into believing that it was self-experienced. HC was not considered a realistic representation of the kind of coaching children may be subjected to in the real world; rather it was the kind of coaching that could be provided only by someone very familiar with the content criteria of CBCA. Thus, this condition was specifically designed to enable an assessment of an important boundary condition of CBCA

evaluation: that is, to determine whether CBCA is sophisticated enough to enable the detection of false statements by children who were provided with the very information needed to fool the system.

3. Lightly Coached (LC), in which children did not experience the event. They were provided with a skeletal outline of the event, with the expectation that it would be up to them to fill in details that would make their reports believable. As in HC, the children were encouraged to imagine their own participation in the event and to attempt to fool the interviewer into believing that their reports reflected self-experienced events. Unlike HC, this coaching was NOT designed to incorporate details meeting CBCA criteria. Thus, LC probably more closely approximated the level of coaching provided in the forensic context to children who are being encouraged by an adult to make a false statement.

Two other tests of the quantitative/qualitative differences between reports by children in the live and coached conditions were included in the present study. The amount and accuracy of detail was compared across the three experimental conditions. As well, a number of qualitative characteristics, suggested by the experimental work of Johnson and her colleagues (e.g., Johnson, 1988) and Schooler and his colleagues (e.g., Schooler, Gerhard, & Loftus, 1986) as distinguishing between real and suggested/imagined memories of adults, were investigated in a preliminary way in the present study.

In the following three chapters, several areas of relevant literature are reviewed. In chapter 2, a brief historical overview of children's treatment as witnesses by the court is provided. The term credibility is then defined according to Undeutsch's (1989) distinction between cognitive and motivational aspects of credibility. Early research (i.e., 1910-1920) on the cognitive aspects of the credibility of children's eyewitness reports is reviewed, and contemporary research (1970's to present) on

children's eyewitnessing abilities is introduced. Undeutsch's distinction between cognitive and motivational aspects of credibility is used to organize the review of contemporary literature on the credibility of children's eyewitness reports presented in Chapters 3 and 4.

In Chapter 3, contemporary research on cognitive aspects of credibility is presented. As well, cognitive-developmental explanations for research findings, and issues of ecological validity relevant to this experimental work are discussed. Chapter 4 is focused on motivational aspects of the credibility of children's reports. The problem of false allegations and the need for a sound method of interviewing children are reviewed. The Step-Wise Interview procedure (the interview protocol used for eliciting children's statements in the present study) is then outlined, followed by a brief discussion of recent methods proposed for assessing the credibility of children's reports. A detailed description of Statement Validity Analysis (including CBCA, the credibility assessment procedure tested in the present study) is provided, followed by a critical review of the small number of field and experimental research investigations of the validity of CBCA. Finally, the need for future research investigating the validity of CBCA is discussed.

In Chapter 5, the present study is introduced, hypotheses are stated, and the methodology is described. Chapters 6 and 7 are comprised of the results of the study and a discussion of these results (respectively).

Chapter 2

INTRODUCTION

There has been a long standing concern about the credibility of children's eyewitness testimony. Although children have been witnesses to, and victims of, criminal acts, they have historically been viewed as extremely unreliable witnesses who are prone to invention, vulnerable to suggestion, and unable to distinguish fact from fantasy (Goodman, 1984a; King & Yuille, 1987). In early canon law, prepubertal children were not permitted to testify under any circumstances. Early British common law was more lenient, allowing children as young as seven years of age to testify on the condition that the court could determine that the child understood the meaning of an oath (Goodman, 1984a). In 1779, the age criterion for admissibility of evidence was dropped. Instead, judges were left to decide on the competence of child witnesses on a case by case basis (Goodman, 1984a, p. 12).

To this date, children's evidence is treated as having special status in the North American and British judicial systems (Davies, Flin, & Baxter, 1986). In Canada, adult witnesses are automatically permitted to testify under oath. In contrast, when a potential witness is under 14 years of age, evidence cannot be heard until the court conducts an inquiry to determine whether the child understands the nature of an oath and is able to communicate the evidence (Canada Evidence Act, Bill C-15). Thus, although less restrictive than preceding government bills, children remain "victims of a discriminatory legal system which developed specific rules regarding

children as inherently unreliable witnesses whose testimony must by specially scrutinized" (Bala, 1989, p. 1). A critical examination of children's eyewitnessing abilities necessitates a review of relevant research literature. In preparation for such a review, the term 'credibility' is defined in section 2.1. Then, the early research on the credibility of children's eyewitness reports is reviewed in section 2.2, and a discussion of factors leading to contemporary research on the credibility of children's eyewitness reports is presented in section 2.3.

2.1 Credibility Defined: Cognitive and Motivational Aspects

In the literature on children's eyewitness testimony, the term credibility is used to refer to two distinct phenomena. First, credibility refers to the accuracy and malleability of children's eyewitness reports. Undeutsch, referring to this as the *cognitive aspect* of credibility, defined it as "the ability to report the details of an observed event accurately and completely..." (Undeutsch, 1989, p. 105). It includes "the eyewitnessing abilities possessed by the individual witness as well as the general factors influencing the acquisition, retention, retrieval and verbal communication of information" (Undeutsch, 1989, p. 105). The majority of research investigations of children's eyewitness testimony have focused on issues related to the cognitive aspect of credibility (e.g., detail and accuracy in reporting, susceptibility to leading questions/information).

The second meaning of the term credibility is related to the truthfulness, per se, of children's testimony. Undeutsch referred

to this as the *motivational aspect* of credibility. He defined it as the individual's willingness to tell the truth about important elements of the crime, his/her own role in the event in question, and the identification of the perpetrator of the event (Undeutsch, 1989).

2.2 Early Research (1910 - 1920)

In the late 1800's, researchers of human sensation, perception, and memory began to deal with issues relevant to the psychology of testimony. A number of writers (e.g., Cattell, 1895; Bolton, 1896; cited in Sporer, 1982) pointed out the potential value of their findings for the criminal justice system. However, reports of systematic investigations of eyewitness testimony did not appear until the turn of the 20th Century. This early work was carried out with predominantly child subjects, and was focused on issues relating the cognitive aspect of credibility.

The earliest reported controlled investigations of children's testimony were conducted by French psychologist Alfred Binet (1900, cited in Goodman, 1984a). Binet explored the susceptibility of school children (aged 7 to 14 years) to suggestion. The children were shown pictures, then were asked for their free recall of what they had seen. As well, they were asked questions varying in the degree to which they were suggestive. The suggestive questioning was reported to have had a dramatic effect on the children's recall. Whereas a number of children correctly reported some facts but not others, there were children who provided detailed false accounts of pictures they had previously seen (Goodman, 1984a).

Binet drew attention to the legal importance of these findings. He suggested that children would be likely to provide more accurate accounts of events if asked to write out their reports rather than being questioned by authorities. However, he recognized the limited generalizability of his findings and called, not only for further research but, for the creation of *a practical science of testimony* (Goodman, 1984a).

In 1901, Stern introduced a research program on the *psychology of testimony* in Germany. In addition to investigations of the eyewitness testimony of adults, he reported on recollection experiments with children. In such experiments, children and young adults (ranging in age from 7 to 10 years) were shown a picture. The picture was subsequently removed and subjects were asked to report on their memory for it. He found that recollection by free narrative resulted in 5 to 10% errors. Testimony given in response to questioning resulted in 25 to 30% errors. Further, he demonstrated a link between age and suggestibility, with leading questions associated with 50% errors in 7 year olds and only 20% errors in 18 year olds (Stern, 1910).

Like Binet, Stern was aware of the limited generalizability of findings based on memory for static pictures. He pioneered the experimental use of *event-tests* in which the to-be-remembered stimuli were simulated live events. On the basis of his findings with adult subjects, Stern (1910) suggested that the reports of witnesses not be assumed to be reliable. As well, he fostered an awareness that "the examining officer is able by the manner of his questioning to predetermine in a measure the degree of the

erroneousness of the testimony. The more he leaves to spontaneous narration, and the less suggestive his questions, the less will be the danger of falsification" (p. 274). He asserted that special consideration be given to the testimony of children and adolescents because "the usual procedure of interrogation greatly diminishes the value of child testimony and at the same time puts the juvenile witness in moral peril" (p. 275). Thus, he suggested the introduction of "special investigating magistrates... before whom the children should be examined but once and then as soon as possible after the event" (p. 275).

In approximately 1904, Stern became the first psychologist to testify regarding the truthfulness of a juvenile witness's testimony in a sexual assault case. On the basis of his review of the successive depositions by the adolescent in question, Stern concluded that the later statements reflected the suggestive questioning used in repeated interviews rather than the youth's recollection of true experiences (Sporer, 1982).

Another early series of studies on children's eyewitness testimony was carried out by the Belgian psychologist Varendonck (1911). This research was conducted against the backdrop of a murder trial in which two young children served as key witnesses. Records revealed that over repeated interrogations, with extensive use of leading questions, the children's eyewitness reports changed considerably. Varendonck was hired by the defense to serve as an expert witness. His task was to discredit the testimony of the two young witnesses.

Varendonck conducted a number of simple investigations designed to demonstrate "how poorly children observe and how suggestible they are" (Varendonck, 1912, cited in Goodman, 1984a, p. 31). In one such study, Varendonck had a teacher ask his class of 7-year-old students to think of another teacher with whom the children were very familiar. The children were then asked to write down the colour of this teacher's beard. Although the teacher in question was in fact beardless, almost all of the children specified a particular colour.

In another study, Varendonck told a class of 8-year-old children that they had witnessed a man approach him in the school playground earlier that morning. During his first interrogation, he suggested to the students that they knew the man in question, and told them to write down his name. Varendonck then simply asked "Wasn't it Monsieur M. who came close to me?", and reminded the children to tell the truth. Although no such event occurred, sixty percent of the students wrote the name of the man suggested by Varendonck. Close to 20 percent gave the name of another individual. Thus, simply by stating the name of the non-existent visitor, Varendonck was able to elicit from the majority of his students reports that they had witnessed an event which had not occurred. On the basis of these findings, Varendonck convinced the courtroom that "the children who testified in the ... [murder] case had seen nothing, absolutely nothing of the murder, nor the murderer; and that consequently, we cannot set the least value in their declarations" (Varendonck, 1912; cited in Goodman, 1984a, p. 31).

Pear and Wyatt (1914, cited in Goodman, 1984a), two British researchers, reported on the first systematic study of children's recall for simulated events. In their experiment, school classes of children (aged 11 to 14) were interrupted by the arrival of two adult strangers. These strangers carried out a staged event in front of the class. The next day, the children were asked to write a free narrative account of the event they had witnessed and then answer a questionnaire asking for specific details of the incident. Pear and Wyatt reported that the free narrative resulted in incomplete but highly accurate reports, with 96% of the statements judged correct. In contrast, 36% of the misleading questions were answered affirmatively, with no discernable difference in the degree of suggestibility evidenced by the two age groups of children (Davies, Flin, & Baxter, 1986).

Consistent with the cautions of Binet, Stern, and Varendonck, Pear and Wyatt stressed the negative effects of suggestive questioning on children's reports. However, with optimism uncharacteristic of the times, they concluded that "the degree of accuracy attained [in spontaneous accounts of events] is remarkably high.... When the testimony of children is unaffected by questions or suggestions, it is worthy of the utmost consideration" (Pear & Wyatt, 1914, cited in Goodman, 1984a, p. 21).

In sum, the findings of the pioneering researchers of children's eyewitness testimony were consistent in drawing attention to the susceptibility of child witnesses to suggestive or misleading information. Although in agreement on this point, the conclusions drawn by these various researchers differed greatly. At one

extreme, Varendonck presented his results as absolute proof of the unreliability of child witnesses. In so doing, he failed to acknowledge that his research was conducted in a remarkably biased manner. Most notably, his research methodology was specifically designed to demonstrate to a courtroom jury the malleability of children's memory. Further, he failed to include adult comparison groups against which to evaluate the suggestibility of children. In contrast, others (e.g., Binet, Stern) noted children's susceptibility to suggestion, but acknowledged the limited generalizability of their findings and encouraged further research to more completely explore children's capabilities as witnesses. Their approach to dealing with the problem of children's apparent suggestibility was far more constructive than was that of Varendonck. Rather than declaring children to be unfit witnesses, these researchers made serious attempts to focus attention on the need for changes in the manner of eliciting testimony from children.

In Europe, most scientific and legal communities treated the early research on children's eyewitness testimony as confirmation of their strongly held belief that children are unreliable witnesses whose evidence is not to be trusted. In Germany, however, the reformatory efforts of, for example, Binet and Stern had some effect on the law of criminal procedures and on the decisions of the Supreme Courts (Sporer, 1982). In 1935, the German court system recommended that psychologist experts be called in to aid the court in evaluating the credibility of child witness' accounts (Undeutsch, 1984).

Developments in North America were different. In 1908, Munsterberg's book *On the Witness Stand* was published. In it, he called upon the legal system to take notice of the findings of eyewitness research, and to reform the American code of criminal procedure accordingly. His fervent advocacy of the value of this psychological research, replete with comments about the ignorance of the legal profession, was received as offensive and condescending by the American legal community. Wigmore (1909), a leading American scholar on evidence, responded with a scathing appraisal of Munsterberg's position. He pointed out limitations of existing eyewitness research and cautioned that the results of such faulty works should not be generalized to the criminal justice system. In accordance with Wigmore's stance, the North American criminal justice system did not accept, or act on, the findings of the early *psychology of testimony* research.

The onset of World War I abruptly ended the early era of research on the *psychology of testimony*. Following the war years, the depressed European economy hindered the return to research productivity, but by the 1930's academic psychology once again began to flourish. This time, though, the major centers of research were based in North America and followed the tradition of behaviourism. Thus, in contrast to the strong pre-war interest in witness testimony and other applied issues, post-war psychology was narrowly focused on behavioural S-R experimentation. The hiatus in research on eyewitness testimony lasted for approximately four decades (Cutshall, 1985).

In the 1960's, Watsonian behaviourism gave way to a renewed appreciation of human cognitive activity. North American academic psychology became interested in issues regarding complex human memory (Yuille & Wells, 1991). By the early 1970's, a new era of applied research on adults' ability to accurately perceive, remember, and report witnessed events emerged. Following this interest in adults' eyewitnessing abilities came a re-emergence of research on children's testimony (Goodman, 1984a).

2.3 Introduction to Contemporary Research (1970's to present)

Renewed research interest in the child witness has, in part, been prompted by findings regarding adults' eyewitnessing abilities. There is an increased awareness of the degree to which adults' testimony may be distorted when they are subjected to misleading information and suggestive questioning (Davies et al., 1986). As well, our present socio-cultural context has played a critical role in returning attention to the credibility of children's testimony. It is now acknowledged that children witness, and are victims of, a variety of criminal acts, including sexual abuse, domestic violence/homicide (Davies et al., 1986), and traffic accidents (Sheehy & Chapman, 1982). However, it is child sexual abuse that most frequently brings children in contact with the criminal justice system. The problem of determining the credibility of child witnesses' reports in sexual abuse cases has played the biggest role in bringing about the current re-evaluation of children's eyewitnessing abilities (Yuille, 1988a).

There has been a dramatic increase in the reporting of child sexual abuse over the past decade. In the U.S., the number of sexually abused children reported to child protection services has increased at a rate of 30 to 35% each year from 1982 to 1984 (Suski, 1986). Finkelhor (1984) estimated that there are 150,000 to 200,000 new cases reported per year in the U.S. Although no comparable incidence figures exist for Canada, it is reasonable to expect that there is a comparable incidence rate, in the neighborhood of 20,000 new cases per year (Yuille, 1988). Because sexual abuse is often kept hidden, official reports of incidence rates are likely gross underestimates of the actual occurrence of such abuse and "may represent only a tip of an unfathomable iceberg" (Finkelhor, 1984, p. 19).

Society has become increasingly concerned with the protection of children's rights. The demand for prosecution of child sexual abuse offenders is mounting. Within this context, evidence provided by children is now regularly heard in both criminal and family courts (Yuille, 1988). Although they have no consensual framework for making such evaluations, the courts are faced with the task of evaluating the testimony (most often uncorroborated) provided by child victims/witnesses. Psychological experts are being called upon for information regarding children's ability to give accurate evidence, and for guidance in making determinations as to the truthfulness of such evidence.

In response to this call for practical information, the experimental literature on children's eyewitness memory is rapidly growing. Children's eyewitnessing abilities have been the topic of

a number of recent conferences/symposia and special edited volumes, including: A 1992 NATO Advanced Study Institute titled *The Child Witness: Psychological, Social, and Legal Perspectives;* a 1989 conference titled *Suggestibility of Child Witnesses* sponsored by the American Psychological Association's Science Directorate, and a resulting text edited by Doris (1991); a symposium on *Adults' Attributions about Child Witnesses* held at the 1987 biennial meeting of the Society for Research in Child Development, and a resulting text edited by Ceci, Ross, and Toglia (1989); a 1985 symposium on *Children's Eyewitness Testimony* and a resulting text edited by Ceci, Toglia, and Ross (1987).

As in the early era of research on children's testimony, most contemporary investigations have focused on assessing children's ability to resist suggestions and give accurate evidence (i.e., the cognitive aspect of credibility). However, we are seeing the growth of a branch of research focused on devising and testing methods for making determinations of the truthfulness of children's evidence (i.e., the motivational aspect of credibility). Advances in our knowledge of both the cognitive and motivational aspects of credibility will be reviewed.

Chapter 3

COGNITIVE ASPECTS OF CREDIBILITY: RESEARCH, THEORY, AND ECOLOGICAL VALIDITY OF STUDIES

3.1 Research Methodology

Typically, investigations of children's testimony have attempted to simulate aspects of real eyewitness situations (Davies, Tarrant, & Flin, 1989). Experimental methodologies used have involved presenting children of varying ages (alone or in groups) with to-be-remembered stimulus events. Events have been presented in many forms. These include: verbal description (e.g., a narrative story about a girl's first day of school, Ceci, Ross & Toglia, 1987; an audiotaped story about a theft, Saywitz, 1987), slide sequence (e.g., of a man stealing a radio, Parker, Haverfield, & Baker Thomas, 1986), film or video (e.g., films depicting episodes of petty crime, Cohen & Harnick, 1980; films of people engaged in social/recreational activities, Dale, Loftus & Rathbun, 1978), and staged event (e.g., a purse snatching, Goetze, 1980; adults arguing, Marin, Holmes, Guth, & Kovac, 1979; a bicycle theft, Yuille, Cutshall, & King, 1988). In all of these studies, children have witnessed the event as a bystander rather than as an active participant in the unfolding drama.

Following delays of varying time intervals, memory for the witnessed event has been tested via free recall, direct questioning, and/or recognition tests. In many of these investigations, a critical issue under investigation has been the influence of misleading information on the accuracy of children's testimony. In

laboratory investigations, suggestibility has most commonly been assessed by presenting misleading information about a witnessed event prior to the recall task, or by subjecting the individual to suggestive questioning during the recall task. The subject's eyewitness testimony has then been examined to assess the degree to which the misleading information has been incorporated into his/her telling of the original event.

3.2 Research Findings

Although the methodologies used in such investigations differed, a number of results have been reported relatively consistently across studies. It has been reported that the amount of information recalled in eyewitness accounts of events varies with age (Yuille, 1988). In free recall of a previously witnessed event, children provided less information than did adults (e.g., Goetze, 1980; Marin et al., 1979). The amount of information provided in free recall appears to increase with increasing age until preadolescence (i.e., age 11 to 12), at which point it reaches adult levels (Cole & Loftus, 1987).

Studies also consistently demonstrated that although children recalled fewer details, their free recall reports were no less accurate than were the reports of older children or adults (Goodman & Helgeson, 1985; King & Yuille, 1987). Children were found to better remember central actions of an event than descriptions of individuals and their physical surroundings (Cole & Loftus, 1987). They were particularly poor at establishing the height, weight, and age of others (Davies, Stephenson-Robb, & Flin, 1988; Yuille,

1988a), and tended to misreport hair/eye colour and clothing details (Davies et al., 1989). There was, however, some evidence for improvement in accuracy of person description from age 7 to 10 (Davies et al., 1989).

Children have been found to give more accurate reports of witnessed events in free recall than when presented with direct questions or recognition tasks (Cole & Loftus, 1987; Davies et al., 1989). Unfortunately, though, the fact that children spontaneously provide relatively little information in their initial, unprompted reports often leads interviewers to ask direct questions of child witnesses. This type of questioning greatly increases the likelihood of inaccurate reporting by the child (e.g., Dent & Stephenson, 1979).

The testimony of adults has been shown to be vulnerable to distortion by presentation of misleading post-event information (Cole & Loftus, 1987; Loftus & Greene, 1980). While there has been much debate over whether children are more suggestible than are adults (see edited volumes by Ceci et al., 1987, and Doris, 1991), results of recent experimental investigations have pointed fairly convincingly to the conclusion that children are in fact more influenced by the presentation of post-event (mis)information than are adults (e.g., Doris, 1991; King & Yuille, 1987; Yuille, 1988). Further, younger children appear more likely than older children to incorporate post-event misinformation into their recall (e.g., Ceci et al., 1987). Yuille (1988) pointed out that children's susceptibility to misinformation is greatest when the children's memory for the original event is poor, the misinformation pertains

to peripheral aspects of the event, and/or the misinformation is presented by a credible source (i.e., an adult).

3.3 Cognitive-Developmental Explanations for Findings

There are a number of explanations for the above findings regarding the cognitive aspects of credibility which draw on the cognitive-developmental literature.

1. Limited attentional capacity. It is well documented that children have a more limited attentional capacity than do adults (Yuille, 1988a). Since younger children are less able to attend simultaneously to multiple aspects of an event, they process and encode into memory less detail about a complex stimulus event that do adults (King & Yuille, 1987). It follows, then, that when later asked for recall of the witnessed stimulus event, children will have less memory to draw on and will provide a smaller amount of information. It has been further demonstrated that when limited by restricted attentional capacity, older individuals focus on central actions rather that peripheral details of an event (Easterbrook, 1959). It may be that children, also limited by a restricted attentional capacity, attend predominantly to salient features such as central actions to the relative exclusion of peripheral details. When later confronted with erroneous post-event information, particularly if it relates to peripheral details of the event (perhaps some physical description of an individual or the physical setting within which action took place), children--with a more limited attentional capacity than adults and therefore less detail

in memory to contradict what is being suggested--will be particularly vulnerable to suggestion (see Doris, 1991).

2. Level of cognitive complexity. Level of cognitive complexity has been posited as a critical determinant of children's less detailed recall and increased susceptibility to suggestion in the eyewitness situation (Yuille, 1988). It is generally believed that the more elaborate an individual's cognitive structures pertaining to an event, the better the individual is able to meaningfully incorporate incoming information about the event into existing knowledge (i.e., encode it into long-term memory), resulting in better subsequent memory for the details of the event (Flavell, 1985; Yuille, 1988a). In most cases, children's cognitive structures relevant to complex eyewitness events are not as rich or elaborate as are the cognitive structures adults bring to bear on such situations. Thus, children have less of a framework on which to hang incoming information. This results in poorer encoding, and a more limited memorial representation of the event (Yuille, 1988a). On this basis, one would expect children's recall of the type of novel events typically presented in eyewitness research to be less detailed than that of adults. Further, with more limited memory for the event and therefore less information to contradict suggestions made, children would be more vulnerable to the effects of post-event suggestions.

Indirect evidence for the role of increased cognitive complexity in making individuals resistant to suggestion can be found in a study by Duncan, Whitney, and Kunen (1982). They

presented children (in Grades 1, 3, and 5) and college students with cartoon slide sequences followed by a series of short answer questions. Each series contained two factual questions asking for information about the slides but not providing any information about the scenarios depicted. As well, the series of questions contained either correct information questions (referring to events depicted in the slides), or misleading information questions (referring to events not depicted in the slides). After a brief delay, subjects were asked follow-up questions designed to assess whether reports of the cartoon events were influenced by the type of questions asked earlier. Based on their finding that performance on the factual questions improved with age, Duncan et al. reported that memory for the visual sequences improved with age. In order to then analyze performance on the follow-up questions independently of the developmental improvement in memory, they assessed performance on follow-up questions only for subjects who correctly answered both of the factual questions presented at the end of the stories. Interestingly, the Grade 1 subjects were unaffected by either correct or misleading information questions, but older subjects (Grades 3 and 5, and college students) were significantly affected by both types of information.

The fact that cartoons served as the to-be-remembered stimuli in this experiment is critical to the interpretation of these findings. It appears that the youngest group of children, being most familiar with and having the most elaborate cognitive structures pertaining to cartoon material, were in this case best

able to resist the effects of post-event (mis)information (Cole & Loftus, 1987).

3. Memory encoding, storage, and retrieval. There is evidence that the mechanisms and skills necessary for the encoding and retrieval of information from memory develop with age (Davies et al., 1989; Raskin & Yuille, 1989). Children, relative to adults, have a more circumscribed repertoire of encoding and retrieval strategies (Zarazoga, 1987). Further, there is recent evidence from studies with subjects ranging in age from late pre-school to early adulthood, that forgetting due to storage failures (i.e., due to the fading of memory traces from long-term memory) does occur, but occurs less often with increasing age (Brainerd & Ornstein, 1991). Thus, when recall for a previously witnessed event is requested, children might be expected to produce a smaller quantity of information than adults because of the more limited amount of information initially encoded into memory, the possibility of trace fading during the retention period, and their inefficiency in systematically and exhaustively searching their memory when attempting to retrieve information. It has been further demonstrated that poor memory for the to-be-remembered event results in heightened suggestibility across age groups (Loftus & Davies, 1984). Perhaps because young children are generally unable to access from memory as many details of the original event as are older children and adults, they are more affected by suggestive questioning (Goodman, 1984b).

4. Children's sensitivity to context. The structural dynamics operating in the recall situation likely play a powerful role in determining children's eyewitness performance (King & Yuille, 1987). Such dynamics have been demonstrated in a number of studies. Yuille, Cutshall, and King (1988) found that when a photo lineup with target (i.e., the individual witnessed carrying out a mock theft) absent was presented to children (age 8 to 14 years), even with instructions informing them that the thief's picture might not be in the lineup, over half of the children falsely identified one of the photos as the witnessed thief. This high rate of false identifications was age-related, with the 8 to 11-year-old children wrongly selecting a photo more often than the 13 to 14-year-olds. King and Yuille (1987) suggested that for young children, the presentation of a photo lineup has much the same effect as presentation of a leading question, eliciting a choice response simply because of the children's inaccurate perception of the task demands. They suggested that prior to the photo-identification task, children be familiarized with the task demands by being presented with mock lineups--one with target (e.g., interviewer's face) present and one with target absent--and be asked to select the interviewer's face in each. Bottoms and Goodman (1989) reported improvement in children's photo-identification performance when such preliminary demonstrations were used to familiarize the children with the task demands prior to the photo-identification test.

Ceci, Ross, and Toglia (1987) found that young children (i.e., preschoolers) were better able to resist the influence of post-event misinformation when it was presented by a child (i.e., a 7-year-old

confederate of the experiment) than when it was presented by an adult. It may be that children incorporate information provided by a credible source into their actual memorial representation of the original event, as was originally suggested by Loftus (1975); however, recent evidence has suggested that this is likely not the case (Lindsay, 1992). Alternatively, children's tendency to regard adults as knowledgeable and competent may lead them to decide that the post-event information provided by an adult is more reliable than their own memory for a witnessed event, and they may therefore choose to report the suggested misinformation instead of their own accurate recall. Similarly, children's desire to please adults may lead them to give testimony in line with the post-event misinformation in spite of their knowledge that the information supplied by the adult is incorrect.

Two additional studies reviewed by King and Yuille (1987) strikingly demonstrate the influence of the dynamics of the situation on children's interview performance. Rose and Blank (1969) conducted a variation on the standard Piagetian conservation task in which they compared the performance of two groups of 6-year-old children. One group was asked for judgments of the equality of two objects before and after one of the objects was manipulated. For example, two identical glasses of water were presented, the child was asked if the glasses contained the same amount of water, water from one of the two glasses was transferred into a glass of a different shape, and the child was again asked to judge whether there were equal amounts of water in the two glasses. The second group was asked for their conservation judgments on only one

occasion (i.e., after manipulation of one of the two objects). In accordance with previous research (e.g., Piaget & Inhelder, 1974), Rose and Blank found that the children in the pre- and post-manipulation questioning condition tended to answer the post-manipulation question incorrectly. However, the children in the post-manipulation questioning only condition were much more likely to give the correct conservation response. Rose and Blank interpreted their findings as evidence that children's failure to display conservation in the standard testing procedure may be due, not to a developmental deficit in ability to judge conservation (as Piaget's theory would suggest) but, to young children's misreading of the task demands. The children may have assumed that if they were being asked the same question again following an obvious experimental manipulation, a different answer was being sought by the interviewer (King & Yuille, 1987).

Hughes and Grieve (1980) presented 5- and 7-year-old children with bizarre, unanswerable questions (e.g., Is red heavier than yellow?). Perhaps surprisingly, instead of admitting their inability to answer such questions, most children of both age groups provided their adult interviewer with 'serious' answers. Hughes and Grieve proposed that much the same process as that involved in children's acquisition of language can be used to explain the children's performance in this interview situation. Specifically, the children actively attempted to derive meaning from unfamiliar input--in this case the bizarre questions--by using available extralingual cues, their knowledge of the world and how objects in the environment are related, etc. Hughes and Grieve concluded that

this demonstrated propensity of children to make sense of and provide answers to bizarre questions necessitates a re-examination of what is being measured by interview procedures aimed at assessing young children's cognitive and linguistic abilities.

In light of the above findings, King and Yuille (1987) suggested the following:

> The term suggestibility can be considered a legal or forensic term for what developmentalists refer to as sensitivity to context.... Context sensitivity is present throughout the life span. For adults it is particularly noticeable when we are dealing with unfamiliar situations.... Relative to adults, children are more suggestible because they find themselves in more situations in which they are unfamiliar.... Thus, children will be more attuned to the social, linguistic, and pragmatic context because it is their means of learning about the world, anticipating appropriate responses, and making the unfamiliar familiar. Consequently, younger children can be expected to be particularly sensitive to contextual cues in a verbal situation where [they are] supposed to listen and respond to questions and instructions from an interviewer. If children are interviewed concerning events they have understood..., and if they are interviewed in a manner that is consistently meaningful and not contradicted by nonverbal cues, then they should be no more suggestible than adults. (p. 30)

5. Reality monitoring. Johnson and her colleagues (see Lindsey & Johnson, 1987) have conducted research addressing the question of whether children's memory for a perceived event is interfered with, not by externally presented misinformation but, by their own internal ruminations and fantasies. They found that children as young as six years of age were generally as able as adults to determine the origin of a memory for an event (i.e., in cases in which they had to discriminate between memories of actions they imagined themselves performing and actions they observed

another person performing). However, children were worse than adults at discriminating between memories for actions they imagined themselves performing and memories for actions they actually carried out. Lindsey and Johnson concluded that this deficit in reality monitoring (i.e., the process involved in discriminating between memories of real and imagined events) "appears to be specific to confusions between self-generated behaviors and imaginings of self-generated behaviors" (p. 107). However, they acknowledged that they have not yet adequately researched the possibility that children may have difficulty discriminating between memories of what they observed another person doing and their memory of what they imagined that same person doing. They drew attention to the importance of this issue in making judgments about the accuracy of children's testimony in a legal context.

In sum, limited attentional capacity, an unsophisticated level of cognitive complexity, a narrow repertoire of strategies for encoding into and retrieving from memory, the fading of memory traces in storage, sensitivity to context, and perhaps failures in reality monitoring, are all factors which may play a role, individually or in interaction, in determining the quantity and quality of children's recall of witnessed events.

3.4 The Problem of Ecological Validity

Although researchers have progressed from the use of non-lifelike stimuli (e.g., stories, slides) to more realistic (albeit staged) events, the ecological validity of much of the experimental

research on children's eyewitness testimony is questionable (Yuille & Wells, 1991). When children have witnessed or been the victim of real-life criminal acts (e.g., sexual abuse), they are asked to provide eyewitness accounts of often traumatic events laden with emotion and personal import. Since it would be ethically unacceptable to intentionally traumatize children in order to later test their memory, most experimental research has assessed children's reports following exposure to emotionally neutral, or at least not personally involving events. Clearly, it would be presumptuous to claim that the cognitive characteristics of children's testimony regarding observed, non-threatening laboratory events would necessarily be comparable to eyewitness accounts following threatening and personally involving real-world events. The ecological validity of research findings must be assessed by comparatively evaluating the social, cognitive, and emotional contexts of the research event and the forensic context to which generalization is intended (Yuille & Wells, 1991).

Generalizability from laboratory to forensic context remains an empirical question that has not yet been adequately tested. Investigations of adult eyewitness memory in real-life crime situations have revealed that testimony provided by adult witnesses to a crime differed in a number of important ways (e.g., accuracy, resistance to suggestion, and persistence of memory over time) from expectations based on laboratory-based investigations of adult eyewitness behavior (Fisher, Geiselman, and Amador, 1989; Yuille & Cutshall, 1986; Yuille & Cutshall, 1989; Yuille & Kim, 1987).

3.5 Recent Research with Improved Ecological Validity

Recently, attempts have been made to address concerns regarding the ecological validity of research findings by examining children's testimony following naturally occurring real-world stressful events. Goodman and her colleagues (e.g., Goodman, Aman, & Hirschman, 1987; Saywitz, Goodman, Nicholas, & Moan, 1991) have conducted a series of studies in which children's recall has been obtained following potentially stressful events which simulate some characteristics of child abuse and investigations of alleged abuse.

Goodman et al. (1987) investigated children's recall for the arousal-producing experience of receiving an inoculation at a medical clinic. After delays of up to nine days, the 3- to 6-year-old children were questioned about their memory for the inoculation. A combination of objective and leading questions about the actions involved, the nurse's physical appearance, and the characteristics of the room were employed. Of note, Goodman et al. included four specific questions deemed likely to be asked in cases of child sexual or physical abuse (i.e., "Did the person kiss you?/ hit you?/ put anything in your mouth?/ touch you anywhere other than your arm or thigh?").

Results indicated that the older children were more accurate than the younger children in response to both objective and suggestive questioning. For children of all ages, objective and suggestive questions about central aspects of the event (i.e., nurse's actions and physical appearance) were answered more accurately than questions about peripheral details (i.e., the room). Responses to the specific questions implying abuse by the nurse were

highly accurate. Thus, Goodman et al. concluded that the "central skein of the action was preserved in even the youngest children and showed resilience in the face of suggestive questioning" (Davies, in press, p. 19).

In a companion study, Saywitz et al. (1991) examined whether children's resistance to abuse related questioning would extend to situations in which the potentially stressful, nonabusive encounter involved the touching of the child's genital area. Saywitz et al. studied the recall performance of girls (aged 5 and 7 years) after a standard medical checkup. The physical examination included an external genital examination for half of the girls, and a scoliosis test for the other half. Children's recall for the checkup was obtained through free recall, re-enactment of the checkup with anatomically-detailed dolls, and specific leading and nonleading questions. Eighty percent of the girls in the genital examination condition failed to spontaneously report the touching and did not demonstrate it with the dolls. However, most (i.e., 31 out of 36) acknowledged it upon direct leading questions. They resisted questioning implying sexual misconduct by the physician. Thus, even when reporting on an interaction that involved touching of their genital area, the children refused to agree with suggestions of abuse. Of interest, though, while none of the girls in the scoliosis condition falsely reported genital touching during their free recall or doll demonstration, 3 out of 36 falsely reported genital touching in response to the specific and leading questions (Goodman & Clarke-Stewart, 1991).

In another experimental study, Rudy and Goodman (1991) involved children (4 and 7 years of age) in a neutral staged event. Pairs of children were led to a trailer, introduced to a male stranger, and left to interact with this stranger. The stranger engaged one child of each pair in a variety of games and had the other child observe. Activities bearing some similarity to descriptions of sexual abuse were included in these games (e.g., tickling, taking photos of the child in different poses). Subsequent eyewitness interviews consisted of open-ended questions, and a number of specific and misleading questions some of which were intended to be suggestive of sexual abuse (e.g., "He took your clothes off, didn't he?", "Did he kiss you?/the other child?"). Free recall for both age groups was highly accurate, with younger children recalling less detail than older children. On specific questions, older children were more accurate and better able to resist suggestive questioning about the confederate and his actions. In general, though, both age groups of children were reportedly "very accurate" in answering questions about potentially abusive actions. Seven-year-olds answered 93% of these questions accurately, and 4-year-olds answered 83% correctly. Most errors were errors of omission which occurred in response to questions about touching, perhaps because of some uncertainty over what constitutes a touch. Thus, Goodman and Clarke-Stewart (1991) concluded the following:

> Children evidenced considerable accuracy in answering specific abuse questions and even in resisting strongly worded suggestions about actions associated with abuse.... These findings counter the view held by many that children are highly suggestible when asked questions about abusive actions. (p. 95)

On the basis of her program of research, Goodman has promoted the view that children are unlikely to falsely report abuse, even when asked suggestive questions implying sexual misconduct by an adult. Such a conclusion is problematic when taken out of the context of the specific investigations from which it was derived. Although Goodman and her colleagues did simulate some of the characteristics of sexual abuse (e.g., touching of genitalia) and resulting interrogations (e.g., "He took your clothes off, didn't he?"), these studies have been criticized for having "taken sexual abuse questions and attached them to events that have nothing to do with sexual abuse" (Peters, cited in DeAngelis, 1989, p. 8). Further, the ecological validity of this research on suggestibility has suffered because children were not provided with any motivation or incentive to present false reports of their experience (Ceci, DeSimone, Putnick, & Nightingale, 1990, cited in Peters, 1991b). As Peters (1991b) pointed out, the motivation in these studies would generally lead children in the direction of providing truthful reports, as "to do otherwise, for example, [to] say they took their clothes off or were sexually touched when they were not, could result in considerable embarrassment for them" (p. 90).

Some recent, as yet unpublished, research (e.g., Bussy, 1990; Ceci, et al., 1990; Warren-Leubecker & Tate, 1990; all cited in Peters, 1991b) has suggested that many children <u>will</u> lie if provided with sufficient motivation to do so. For example, Ceci et al. (1990, cited in Ceci, 1991) conducted a study in which, prior to being interviewed, preschoolers were informed by the interviewer that it is naughty for an adult to kiss a child when the child is

naked. In their later statements, the children falsely reported that they were not kissed while being bathed (even by their own parent). Thus, when provided with incentive to withhold the truth, in this case incentive similar to that thought to be operating in cases of alleged sexual abuse (e.g., desire to protect the adult, fear of embarrassment, fear of negative reaction of interviewer and/or the accused adult, etc.), children gave false reports. Ceci and his colleagues are continuing to examine the effects of various types of motivational inducements on errors of omission and commission in children's statements (Ceci, 1991).

Peters (1987, 1991a) conducted investigations examining the effect of different levels of stress on children's recall. Rather than pursuing abuse related themes (as did Goodman), children were asked questions arising directly out of the experienced events (Davies et al., 1989). Peters (1987) tested the eyewitness memory of children (aged 3 to 8 years) for a visit to the dentist. The children's recognition memory for various aspects of the visit (e.g., memory for dentist's face, voice, and examining room) was assessed following retention intervals of various lengths. Although Peters innovatively attempted to use a naturally occurring stressful event as the to-be-remembered stimulus event, methodological problems limited the internal validity of the study and consequently the generalizability of the obtained results. Specifically, the level of stress experienced by the children during dental visits was far lower than expected, thus potentially accounting for the absence of many hypothesized stress effects, and reducing the applicability

of the findings of this study to real-world forensic contexts in which children have been traumatized by crime.

In a later investigation, Peters (1991a) manipulated children's level of stress during a staged event and examined the effects of stress and misleading information on children's later eyewitness performance. Children (6 to 9 years of age) were individually brought into the laboratory, fitted with blood pressure and pulse monitoring equipment, and engaged in a card game with the experimenter. For children in the high stress condition, the game was interrupted by the sounding of a fire alarm and the entry of a confederate who reported that she smelled smoke and expressed concern about a possible fire. Approximately 1 minute later, following the confederate's departure, the alarm was turned off and the children were reassured that there was no fire. Children in the low stress group heard, instead of an alarm, a radio being turned on. The confederate entered the room and talked with the experimenter without mentioning smoke. As in the high stress condition, she departed about 1 minute later. Heart rate and pulse data made it possible to check the effect of the stress manipulation. In fact, the manipulation was effective. All children were then asked a series of questions, some of which contained misinformation. After a delay, their recall for the event was tested with forced choice questions designed to determine the extent to which their recall incorporated the misleading information.

Children in the high stress group were found to perform modestly but significantly less well than children in the low

stress group on both objective and misleading questions. Thus, Peters' results suggest that children's recall of an event is negatively affected by higher levels of stress.

Generalizability of these findings to the forensic context will likely depend on the degree to which the manipulated level of stress during the staged event is representative of the level of stress experienced by children in forensically relevant eyewitness situations. Further, as Peters (1991a) and Goodman (1991) have agreed, the practical applications of these research findings will depend on the extent to which the social and motivational contexts within which children's eyewitness reports are obtained approximate the social and motivational contexts that children find themselves in when giving testimony in criminal cases. Finally, the ecological validity of the study can be expected to depend on the degree to which the type of questions used in this study reflect the questioning carried out in interrogations of children who have witnessed or been victims of real-life crime. Davies et al. (1989) pointed out that in forensically relevant abuse cases, the event in question is often not as clearly structured, nor is the questioning comprised of objective questions from a neutral interviewer, as was the case in Peters' recent laboratory investigation.

Clarke-Stewart, Thompson, and Lepore (1989) reported on an investigation in which they attempted to study children's recall in a situation in which the event contained some interpretive ambiguity and later questioning was anything but objective. Children (5 to 6 years of age) in two experimental conditions were individually exposed to a staged event. The specific actions of the adult

confederate involved in the event were balanced across groups, but the script followed and the interpretations placed on his actions differed. In one scenario, the confederate (Chester the janitor) tidied the room, then cleaned and arranged some toys. He specifically attended to a particular doll, spraying and wiping its face, looking under its clothes for dirt, rearranging its limbs, and biting off a loose thread. While doing so, he made comments consistent with having the aim of cleaning the doll. In the second scenario, Chester handled the doll in a rougher manner and made suggestive comments, to the effect that he likes to "play with dolls,...spray them in the face,...look under their clothes... bite and twist their arms and legs" (Clarke-Stewart et al., p. 2).

After a delay of approximately 1 hour, an adult interviewer posing as Chester's boss questioned each child about Chester's actions. Children were randomly assigned to interrogations that were neutral, incriminating (i.e., accusing Chester of playing with the toys), or exculpating of his actions (i.e., suggesting that he was only cleaning the toys). The children in the incriminating and exculpating interrogation conditions were interviewed a second time (by a second interviewer) using either an interrogation style identical to the first, or switching to the opposite type of questioning.

Children in the neutral interrogation condition gave limited but accurate responses to open-ended questions, and answered over 80% of the specific factual questions and interpretive questions accurately. They continued to answer questions accurately on follow-up one week later. The picture was strikingly different for

children in the biased interrogation conditions. One quarter of the children who were subjected to interrogations inconsistent with what they actually saw (e.g., saw Chester clean, interrogated by interviewer accusing Chester of playing) answered interpretive questions inaccurately after the interviewer's first gentle suggestion. Less than half of the children maintained their original accurate interpretation of the event over the course of the interrogation (which consisted of progressively stronger suggestions). One third totally switched their interpretation of events to be consistent with the interrogator's suggestions. The remaining one quarter of the children incorporated elements of the suggestions into their answers, stating that Chester both cleaned and played.

When the second interrogation was of the same type as the first, children continued to answer in line with the suggestions planted in the first interview. Even upon questioning by their own parents, and at the 1 week follow-up, children continued to answer questions in a manner consistent with the interviewer's interpretation of Chester's activities. When the second interrogation was contradictory to the first, the children generally switched interpretations. One week later, they remembered both interviewers' interpretations and incorporated both into their own version of what happened.

Generalizability of Clarke-Stewart et al.'s findings to the forensic context is a matter for further study. However, through attention to a variety of factors enhancing the ecological validity of their study (e.g., event with interpretive ambiguity, repeated

and varied interrogations), Clarke-Stewart et al. were able to shed some light on the issue of why children change their stories to go along with suggestive questioning.

The finding that children subjected to the suggestive interrogation responded to the 17 factual questions about Chester's behavior as accurately as children in the neutral condition argues against the possibility that the children changed their stories because their memory for the facts of the event were distorted by the suggestive questioning. Further, the finding that children maintained an interpretation of the event consistent with the interrogator's suggestive questioning even when later questioned by their own parents argues against a demand characteristics interpretation of the data. If children were only going along with the suggestive questioning in order to please the adult authority, there would be no need to maintain the distorted interpretation upon later questioning by their own unbiased parents.

Clarke-Stewart et al. argued that the data support the hypothesis that when faced with an interpretively ambiguous situation (or one which they do not clearly understand), children will accept as true the adult interrogator's interpretation of the event. As support for this interpretation, Clarke-Stewart et al. pointed out that although children in this study maintained accurate recall for the facts of the event, they changed their interpretation of the event to match the adult interrogator's view. Further, the fact that at 1 week follow-up, those children whose two interrogations were contradictory remembered both interpretations and combined them in their stated personal interpretation of the

event seems to support the idea that children were using the adult's perspective to guide their understanding of the situation.

In conclusion, Clarke-Stewart et al. pointed out the "very real risks of suggestive interrogation of child witnesses and ... the ease with which interviewers can bias children's interpretation of unusual events they have observed or participated in" (p. 7). This work demonstrated the importance of ensuring that children are interviewed in an unbiased manner by interviewers with no vested interest in any particular interpretation of the events in question.

3.6 Summary of Contemporary Research Findings Regarding Cognitive Aspects of Credibility

The accumulated research to date on cognitive aspects of credibility has greatly improved our understanding of the accuracy and malleability of children's memory for experimental eyewitness events. Younger children have been found to provide a smaller quantity of detail and to be more vulnerable to the effects of misleading information than are older children and adults. Nevertheless, studies have convincingly demonstrated that when recall is solicited in an unbiased manner, and by a neutral interviewer, children are capable of providing reasonably accurate accounts of witnessed events.

Inaccuracies or distortions in reporting can occur for reasons other than leading interview techniques. The above research findings regarding children's cognitive capabilities as witnesses are based largely on investigations in which children were provided with little in the way of motives (e.g., threats, inducements,

suggestions, rewards, etc., Ceci, 1991) to falsely report or fabricate their eyewitness reports. In the forensic context, while children may be capable of providing accurate testimony, cases of deliberate distortion or falsification of reports do occur. To date, little research has been conducted to specifically address the problem of assessing the truthfulness of testimony. In the following chapter, the literature and research findings regarding motivational aspects of credibility will be reviewed.

Chapter 4

MOTIVATIONAL ASPECTS OF CREDIBILITY: FALSE ALLEGATIONS, ASSESSMENT METHODS, RESEARCH, AND FUTURE DIRECTIONS

For years, assessments of motivation to tell the truth have been made on the basis of evidence regarding the "character and conduct of the witness" (Undeutsch, 1989, p. 107). The judiciary has been expected to evaluate the credibility of evidence on the basis of the witness's general reputation for truthfulness as well as specific instances of conduct reflective of honesty or dishonesty. Although such evidence continues to be admitted in North American courts of law, research findings have cast grave doubts on the value of character evidence for determining the truthfulness of a particular statement. Undeutsch (1989) pointed out that such evidence is unreliable and does not take into account the motivational set of the reporting individual. As an alternative, he proposed that the assessment of an individual's motivation to tell the truth be conducted, not by evaluation of the character of the reporting individual but, by evaluating the truthfulness of the individual's statement itself.

At a theoretical level, it is interesting to postulate the qualities critical for enabling distinctions to be drawn between children's statements based on memory for actual experiences (i.e., truthful statements) and those that have been fabricated or based on suggestion/coaching by others (i.e., false statements). At a very practical level, the need for bettering our understanding of how to

make such distinctions is paramount in the area of child sexual abuse.

In section 4.1, estimated rates of false allegations of child sexual abuse are reported. Then, in sections 4.2, the need for a sound interview procedure is outlined and the Step-Wise Interview procedure is briefly described. As well, section 4.2 highlights the need for a valid method of assessing the credibility of children's eyewitness reports, briefly reviews experimental research aimed at identifying features distinguishing between real and suggested memories, and sets the stage for the introduction of the most developed credibility assessment procedure to date: Statement Validity Analysis (SVA).

4.1 The Problem of False Allegations

While figures regarding rates of true versus false allegations of child sexual abuse must be viewed as estimates only, a number of researchers have endeavored to investigate the frequency of such reports. In the most methodologically sound incidence study to date, Jones and McGraw (1987) studied the 576 cases of child sexual abuse reported to the Denver Department of Social Services during 1983. Using their own validation procedure, 53% of all cases were judged to be 'founded' reports. The rest, 47%, were classified as 'unfounded'. The researchers reported that approximately one-half of the unfounded cases were labelled as such because of insufficient information to allow a conclusion regarding whether or not child sexual abuse had occurred. Thus, the 47% of cases deemed unfounded may, because of inadequacies in the validation procedures used, have

included a number of true allegations. In fact, Jones and McGraw reported anecdotally that a number of the cases judged to be unfounded on the basis of insufficient information surfaced as confirmed cases within the two years following their incidence study. Seven percent (i.e., 34 cases) of the total 576 reported cases were judged to be false allegations. Of these, most (i.e., 26 cases) were brought forward by adults on behalf of their children.

Recent reports have suggested that the rate of false disclosures of abuse is much higher in the context of custody and visitation disputes. Jones and Seig (1987) estimated a false disclosure rate of 28% in such a sample, a rate four times that found in the earlier general incidence study (Jones & McGraw, 1987). Indications are that allegations of sexual abuse are being raised with increasing frequency in parental disputes over custody and access (see Benedek & Schetky, 1985; Green, 1986; Sink, 1987).

Children who falsely report abuse may be doing so with the encouragement of a vindictive parent, who in the heat of a custody dispute is using the child as a weapon to gain an advantage over the other parent (Gordon, 1985). Some children may be pressured into accepting as true an over-anxious parent's mis-perception of the relationship between the child and the alleged abuser. In other cases, the false allegation may be made by the child as an expression of anger or a call for help (Yuille, Hunter, & Harvey, 1990), or because of an honest misinterpretation of non-abusive behavior primed by the current focus on teaching children to be vigilant for potential sexual abuse. Cases have also been noted in

which fictitious accounts have resulted from accusations improperly elicited by professional interviewers (Jones & McGraw, 1987).

"It is sobering to realize that we have no reliable, validated methods for identifying and confirming credible cases of child molestation, nor can we readily discriminate instances of false or unreliable allegations" (Rogers, 1990, p. 57). As a consequence, it has been estimated that with a false allegation rate of approximately 8%, there could be over 8,000 serious legal actions and false prosecutions in the U.S. in one year alone (Raskin & Yuille, 1989). Although there are presently no comparable data on the rate of sexual abuse or of false allegations in Canada, there is reason to believe it may be following the same pattern as that in the U.S. (Yuille, 1988a).

In the present social and legal environment regarding child sexual abuse, misguided prosecution of false allegations brings harm not only to the wrongly accused, but to the child involved as well. The pressures, stresses, and emotional conflicts which lead to the initial false disclosure, and are likely exacerbated throughout the legal proceedings, threaten the emotional health of any child entangled in a false allegation (Yuille, Hunter, & Harvey, 1990).

The increasing number of cases of false disclosures (see Green, 1986; Sink, 1987) may have a detrimental effect on children with credible reports of abuse as well. The frequency with which false allegations proceed to formal actions runs the risk of priming triers of fact to the possibility of fabricated allegations, thus potentially biasing them against accepting as credible a child's true report of abuse (Yuille, 1988a). When this occurs, children

are victimized not only by the initial abuse, but by the secondary trauma resulting from their mistreatment by the legal system (Bala, 1989). Further, the accused adult is freed to continue his/her abuse of children.

As Yuille (1988a) suggested, "the more effective we can be at identifying the minority of false disclosures by children, the more confident we can be in supporting judicial changes which give more credence to the testimony of children" (p. 259). In order to become more effective at discriminating true from false allegations, Yuille and his colleagues (e.g., Raskin & Yuille, 1989; Yuille, Hunter, Joffe, & Zaparniuk, in press) have stressed the importance of the interview procedure used, and the need for a systematic procedure for assessing the credibility of children's statements.

4.2 Need for Interview and Credibility Assessment Procedure

Interview

At present, many of the interview methods commonly used for eliciting children's testimony are far from adequate. The tremendous increase in reporting of sexual abuse has brought about an unsatisfactory situation in which children are interviewed, in many cases repeatedly, by professionals who approach the interview with biases and preconceptions, are uncomfortable discussing sexual matters, undereducated in the dynamics of sexual abuse, insensitive to issues regarding children's level of cognitive and language development, and inadequately trained to deal with the special considerations in interviewing children (see Gelfand & Raskin, 1988;

Raskin & Yuille, 1989). There is an urgent need for a systematized interview procedure that can be used to maximize the amount and accuracy of information obtained by child witnesses while minimizing the bias or distortion introduced by the questioning.

The Step-Wise Interview procedure, used to elicit the eyewitness reports of children in the present study, was introduced by Yuille and his colleagues (e.g., Yuille et al., in press) to meet this need. The purpose of the Step-Wise Interview is to obtain from the child as extensive a statement as possible without in any way leading or biasing the child's report. The entire interview is videotaped and/or audiotaped, then transcribed for later statement analysis using CBCA. The interview is divided into four distinct phases which are described below.

1. Rapport Building. The interviewer asks the child neutral questions (e.g., favourite subject at school) in an effort to put the child at ease. As well, the child is asked to describe one or two personally experienced events unrelated to the event that is the focus of the interview (e.g., a birthday party or school outing). The child's general level of linguistic, cognitive, behavioral and social skills are observed during this interaction for comparison with later behavior when discussing the event of concern. The meaning of truth and deception is discussed, and the importance of telling the truth is emphasized. The purpose of the interview is introduced using open questions (e.g., "Do you know why you are talking to me today?"). Only if these type of questions do not bring the topic into the open, the interviewer proceeds to more specific questions (e.g., "Has anything happened to you which you would like to tell me about?").

2. Free Recall. When rapport has been established, the free recall phase begins. The aim of this phase is to provide the child with every opportunity to disclose his/her own account of the events. The child is asked to start at the beginning and describe everything (s)he remembers about the events in question. The interviewer does not interrupt, and certainly does not correct or challenge, the child's report. The child is allowed to proceed at his/her own rate. Thus, patience and a tolerance for pauses and elaborations on potentially irrelevant detail are critical. If the child becomes silent,

the interviewer encourages him/her to continue by asking open ended, non-leading questions (e.g., "and then what happened?").

3. Open-ended Questions. In the third phase of the interview, open-ended questions are asked in order to obtain elaboration of details described in the earlier free narrative. Special care is taken to ensure that the questions are not leading. Further, questions are phrased in a manner that implies that an inability to recall the detail in question is acceptable (e.g., "Is there anything else you remember about _____?"). An attempt is made to separate memory difficulties from a reluctance to talk about certain topics by suggesting that the child use a signal (e.g., raised hand) when (s)he does not feel ready to talk about a given topic. The topic is then raised again later in the interview.

4. Specific Questions. This phase is included in order to allow for clarification and extension of previous answers. The interviewer ensures that questions asked do not include information obtained from other individuals, and makes an effort to avoid providing alternative answers when asking a question. The origins of language/knowledge displayed that seem inappropriate for the child's age are explored. Inconsistencies in the child's statement are addressed with gently probing questions (e.g., "You said it happened when you just woke up in the morning but you said you had your boots on. Can you tell me how that happened?"). After specific questions have been asked, the interviewer asks the child to once again describe the events in question (or some part of the narrative). In asking for this repetition, the interviewer makes clear that (s)he is attempting to understand the event and is not doubting the child's story.

Credibility Assessment Procedure

The task of devising a sound system for determining credibility is a difficult one. In recent years, a number of practitioners have attempted to use their clinical expertise to develop models and strategies for assessing the validity of children's statements in sexual abuse cases (e.g., de Young, 1986; Garbarino, Guttman, & Seeley, 1986, cited in Gelfand & Raskin, 1988; Gardner, 1987; Green, 1986; MacFarlane & Krebs, 1986; Quinn, 1988). The products of these efforts have been uniformly disappointing, and

in some cases can be considered "potentially dangerous documents" (Gelfand & Raskin, 1988, p. 28). Some authors have listed a variety of issues thought to be important considerations in determining when children are engaging in deception (e.g., Quinn, 1988). Others enthusiastically promote a clinical decision model (e.g., Green, 1986) or questionnaire (e.g., The Sexual Abuse Legitimacy Questionnaire, Gardner, 1987) based on "zeal and personal conviction" rather than on scientific standards of research-based validation (Gelfand & Raskin, 1988, p. 28).

Research findings from investigators in the cognitive-experimental tradition appear to have much to offer to the development of a method of assessing credibility. Ekman's work (see Ekman, 1985) on nonverbal clues to deception is clearly worthy of attention. Through his research, Ekman has gathered information on the behaviors (e.g., facial expressions, body movements, voice, words) that may provide clues to when an individual is deliberately attempting to mislead. These findings have yet to be tested in the forensic context, and have not yet been applied to children. Such applications, though, are beyond the scope of the present thesis.

The experimental work of Schooler and his colleagues (Schooler, Gerhard, and Loftus, 1986; Schooler, Clark, & Loftus, 1988), and Johnson and her colleagues (e.g., Johnson & Raye, 1981) could play a role in guiding the science of credibility assessment. Schooler et al. (1986) conducted an experimental investigation of the qualitative differences between real and suggested memories. Two groups of adult subjects witnessed a slide sequence depicting a traffic accident. For one group, the sequence included a slide with

a yield sign present. For the other group, the presence of the sign was merely suggested. Subjects' later recall for the sign was obtained. Compared to real memory descriptions, descriptions based on suggestion were found to be longer, and to contain more verbal hedges, more references to cognitive operations, more self-references, and fewer sensory details (Schooler et al., 1988). These results were replicated with a different stimulus object. Interestingly, untrained judges (undergraduates) who were presented with the transcribed memory descriptions were found to have a limited ability to distinguish real from suggested memories. However, when provided with information regarding the previously determined qualitative differences between real and suggested memories, their classification decisions improved greatly. Schooler et al. concluded that "it may be possible to develop a set of generic hints that can help people to more accurately determine the source of a memory" (p. 179).

The Reality Monitoring research carried out by Johnson and her colleagues (e.g., Johnson, Foley, Suengas, & Raye, 1988; Johnson & Suengas, 1989; Suengas & Johnson, 1988) has focused primarily on studying the characteristics used by adults in distinguishing between their own autobiographical memories for perceived and imagined events. Through this work, a number of qualitative characteristics differentiating real and imagined memories have been identified. Compared with memories for imagined events, subjects' memories for perceived events were longer (the opposite pattern to what Schooler et al., 1988, reported), contained more contextual (i.e., temporal and spatial) information, more sensory details, and

were more likely to give rise to supporting memories (Johnson et al., 1988). Johnson has since begun to examine peoples' ability to use these distinguishing features in making judgments about the origins of others' memories (Johnson & Suengas, 1989). This research has potentially exciting applications to the area of witness testimony evaluations.

The qualitative characteristics identified by Johnson and her colleagues as differentiating memories for real versus imagined events were derived from empirical investigations aimed at systematically examining qualitative characteristics of mental experience (Johnson, 1988). Interestingly, the qualitative criteria identified through this program of scientifically rigorous research are reminiscent of a number of *criteria of reality* identified by Undeutsch (1954, p. 146, cited in Undeutsch, 1984) as important in determining the credibility of children's statements. Unlike Johnson et al., though, Undeutsch arrived at these qualitative criteria on the basis of his practical experience serving as expert witness to the German courts in criminal cases involving child witnesses. The statement analysis procedure pioneered by Undeutsch is the most developed system to date for assessing the credibility of children's testimony (Davies, in press). In sections 4.3 to 4.8, this statement analysis procedure, now known as Statement Validity Analysis, is described.

4.3 Statement Validity Analysis (SVA)

4.4 Historical Origins

In Germany, a 1954 Supreme Court decision suggested that an expert psychologist or psychiatrist be used to aid the courts in the evaluation of the credibility of child testimony in criminal cases, particularly sex cases, in which uncorroborated evidence provided by a child is central to the criminal proceedings (Undeutsch, 1989). As a result, psychologists in Germany have offered expert opinion regarding the credibility of testimony in approximately 40,000 cases during the years between 1950 and 1980 (Arntzen, 1982, cited in Undeutsch, 1989). On the basis of his experience as expert witness in cases of child sexual abuse, Undeutsch developed the first statement analysis approach to the evaluation of children's evidence.

The basic assumption of Undeutsch's statement analysis approach is that "statements which are based on observation of real (self-experienced) events are different in quality from statements which are not based on observations but are mere products of fantasy [including fabrication or coaching]" (Stellar & Koehnken, 1990, p. 3). Undeutsch developed a set of reality criteria that he proposed reflect specific features differentiating truthful from invented testimony (Undeutsch, 1989).

This general approach, known as Statement Reality Analysis (SRA), was further developed over the years by Arntzen (1970) in West Germany, and Szewzcyk (1973) in East Germany (Undeutsch, 1989). SRA has been used routinely by forensic psychologists in Germany and

Sweden since at least 1968 (Undeutsch, 1989). The technique has shown promise as a means of discriminating valid from invalid statements (see Undeutsch, 1982, p. 49). However, proponents of SRA, claiming that its use requires sophisticated clinical skill, have not explicitly detailed the procedures involved in a way that would allow its reliability and validity to be empirically investigated. Thus, it would seem that Undeutsch (1989) was hasty in claiming that "it turned out that this approach of assessing the truthfulness of testimony statements is superior to a common sense evaluation of witness evidence in ... proving the veracity of some statements and in revealing the unreliability of other statements" (p. 116).

An international contingent of psychologists endeavored to systematically modify SRA by providing more specific descriptions of the criteria of reality in order to make the procedure amenable to empirical investigation. This effort by Stellar and Koehnken of University of Kiel, West Germany; Raskin of University of Utah; and Yuille, of University of British Columbia, resulted in what is now referred to as Statement Validity Analysis (SVA).

4.5 General Description

SVA is comprised of two major components: a statement analysis procedure referred to as Criteria-Based Content Analysis (CBCA), and a Validity Checklist. Much of the following material outlining CBCA and the Validity Checklist has been taken from the writings of Yuille and colleagues (e.g., Yuille, 1990a,b,c; Yuille, 1988a; Yuille & Farr, 1987; Raskin & Yuille, 1989).

4.6 Criteria-Based Content Analysis (CBCA). The child's statement is evaluated with respect to 19 content criteria which are grouped into five major categories (See Table 1). The content criteria are described as follows.

I. Criteria Relating to the General Characteristics of the Statement.

This first step of the statement analysis procedure involves examining the children's testimony as a whole. The formal structure of the statement is evaluated in terms of the three characteristics:

1. Coherence. This refers to the degree to which the statement is homogeneous, with details that fit together to form a coherent and internally/logically consistent account of events.

2. Spontaneous Reproduction. A statement is judged to be more credible if it is provided in a spontaneous and somewhat disorganized fashion rather than in rigid form and in perfect chronological order. The difference in spontaneity between a credible and noncredible account is likely to be accentuated upon later retelling of the events. On repetition, the credible report will likely be presented in somewhat different form than on the first telling (e.g., additions or deletions of peripheral detail, different sequence of reproduction), whereas the noncredible account will likely more precisely imitate the initial recounting of events.

3. Sufficient Detail. With developmental differences in children's eyewitnessing ability taken into account, the greater the amount of detail (e.g., elaborate description of person, place, event, and peripheral detail) reported, the more likely the report is judged to be credible. Noncredible statements are "usually impoverished in specific details, and [interviewer prompting] ... will usually elicit few additional details" (Yuille, 1990b, p. 3).

II. Specific Contents of the Statement.

In this second stage of the content analysis, the particular contents of the statement are evaluated with respect to the following characteristics:

4. Contextual Embedding. The telling of an event within its spatial and temporal context is viewed as enhancing credibility. Although the context within which some

events occur may not be memorable, a truly experienced event is more likely than a fictional account of an event to be described in relation to the place, time, and interactions/occurrences that surrounded it.

5. Description of Interactions. This refers to the reporting of a chain of actions and reactions (e.g., conversation, behavior) between the child witness and the other individual(s) involved in the event. This criterion is better met when the child's report is elaborate enough to describe the flow of activities during the event (Yuille, 1990b).

6. Reproduction of Conversation. Verbatim reproduction of dialogue (as opposed to recounting of the general content of conversation, which would support Criterion 5) would fulfill this criterion. Credibility is enhanced when the child's reproduction of dialogue includes vocabulary atypical of the child's age, the child quotes arguments made by the adult, or the conversation reveals the differing attitudes of the adult and child.

7. Unexpected Complications During the Incident. This criterion is fulfilled when the child reports either a spontaneous termination to the event or a chance happening (e.g., knock at door) which interrupts the incident.

III. Peculiarities of the Content.

8. Unusual Details. This refers to the inclusion of detail that has a low probability of occurrence yet is not completely unrealistic. Mention of an unusual item or a common item used in an unusual way would satisfy this criterion. It is assumed that a child presenting a fabricated or coached account is unlikely to incorporate detail with such a low probability of occurrence.

9. Peripheral Details. This refers to the inclusion of concrete and vivid descriptions of details that, are not unusual but, are not pertinent to the central aspects of the event. Similar to Criterion 8, it is thought that children are unlikely to be sufficiently sophisticated at fabricating to include such seemingly irrelevant detail.

10. Accurately Reported Details Not Understood. This refers to the child's reporting of an event that (s)he does not understand, but nevertheless presents in a clear manner that can be understood by the adult interviewer. "The occurrence of this criterion in a statement is supported if a child witness falsely interprets a (correctly described) observation (Stellar & Koehnken, 1990, p. 13).

Table 1

Criteria-Based Content Analysis: Summary of the 19 criteria

GENERAL CHARACTERISTICS OF THE STATEMENT

1. Coherence
2. Spontaneous Reproduction
3. Sufficient Detail

SPECIFIC CONTENTS OF THE STATEMENT

4. Contextual Embedding
5. Descriptions of Interactions
6. Reproduction of Conversation
7. Unexpected Complications During the Incident

PECULIARITIES OF THE CONTENT

8. Unusual Details
9. Peripheral Details
10. Accurately Reported Details Not Understood
11. Related External Associations
12. Accounts of Subjective Mental State
13. Attribution of Perpetrator's Mental State

MOTIVATION-RELATED CONTENTS

14. Spontaneous Corrections
15. Admitting Lack of Memory
16. Raising Doubts About One's Own Testimony
17. Self-deprecation
18. Pardoning the Perpetrator

OFFENSE-SPECIFIC ELEMENTS

19. Details Characteristic of the Offense

11. Related External Associations. This criterion is fulfilled when the child makes reference to events or relationships external to the immediate events being described. Such a reference is at least tangentially related to the key incident but is not integral to it. For example, a child alleging sexual abuse may report that the accused asked him/her to describe the extent of his/her previous sexual experience.

12. Accounts of Subjective Mental State. This refers to the child's accounts of his/her own cognitive and emotional states at the time of the events being described. The description of changes in emotion or cognition during the course of the event enhances the fulfillment of this criterion (Stellar & Koehnken, 1990).

13. Attribution of Perpetrator's Mental State. Comments reflecting inferences about the cognitive and/or emotional state of the adult involved in the event are considered to be credibility enhancing.

IV. Motivation-Related Contents.

The criteria in this category are inferred from the content of a transcribed statement. They play a role in allowing an assessment of the witness' motivation to provide false testimony (Stellar & Koehnken, 1990). Criteria in this category are all considered very unlikely to occur in fabricated or coached accounts because the unsophisticated child is likely to view such admissions as detracting from the believability of his/her report. These criteria include the following:

14. Spontaneous Corrections. When a child spontaneously corrects him/herself during the interview, this is seen as enhancing credibility, particularly if the correction reflects a new clearer recollection. Corrections which take place in reaction to the interviewer's questions or suggestions are not considered spontaneous corrections.

15. Admitting Lack of Memory. This criterion is fulfilled when a child indicates, either spontaneously or in response to a question, that (s)he does not remember certain details of the event. Such a spontaneous admission is unlikely to occur in a false statement.

16. Raising Doubts About One's Own Testimony. When a child "expresses objections to the correctness of [his/her] own testimony" (Undeutsch, 1967, p. 153, cited in Yuille, 1990b), credibility is enhanced.

17. Self-Deprecation. Inclusion of unfavourable self-incriminating details (e.g., mention of a supposed wrong behavior toward the adult subject of the testimony) is considered to fulfill this criterion.

18. Pardoning the Perpetrator. Providing explanations or rationalizations for the accused's behavior fulfills this criterion. It is viewed as credibility enhancing because children intent on blaming the identified suspect are unlikely to make efforts to exonerate him/her.

V. Offense-Specific Elements.

19. Details Characteristic of the Offense. This criterion is probably the most closely tied to the circumstances of sexual abuse. It is fulfilled when the child's description of the course of events fit with what is known about the typical ways in which sexual abuse of children develops.

In the present study, the criteria were slightly modified. Criterion 19, Details Characteristic of the Offense, was simply not applicable to eyewitness reports based on the relatively innocuous staged event. Thus, it was dropped. Two additional criteria were assessed on an experimental basis. In the pilot testing, it became clear that descriptions by child witnesses of their observations of the adult confederate's behavior that was not part of an interaction, or of their own behavior that was not part of an interaction, were being overlooked by CBCA evaluation. These actions were often reported in a vivid manner and seemed to reflect credibility. Thus, CBCA evaluators noted them as follows: Criterion 20, Reports of Other's Action, referring to actions of the confederate actor that did not occur in the context of an interaction with the child witness (e.g., "then he walked over to the lamp and started unscrewing off the top"), and Criterion 21, Reports of Own Action, referring to actions of the child that did

not occur as part of an interaction with the confederate (e.g., "I kept building the LEGO house").

4.7 Validity Checklist. Following an assessment of the contents of the child's statement, the expert evaluates other sources of information in order to make a judgment regarding the credibility of the child's report. This evaluation includes an assessment of four general areas.

(a) The Child's Behavior. Were language, affect, and gestures appropriate to the situation? How susceptible was the child to suggestion? What kind/level of sexual knowledge was displayed (e.g., in verbal reports, drawings, and behavior with dolls)? Was there evidence of sexualized behavior towards him/herself or towards the interviewer?

(b) Interview Characteristics. Was the interview conducted appropriately, with adequate establishment of rapport and opportunity for the child to give his/her free narrative account of the events? Were suggestive/leading questions asked or pressure/coersion applied? If so, were these factors present to the extent that they would compromise the use of SVA?

(c) Motivational Considerations. The context of the original disclosure is evaluated. For example, was the disclosure spontaneously initiated by the child? If initiated by the child, did he/she have reason to report abuse to achieve some end (e.g., removal from present living situation)? Was the disclosure initiated by a parent? If so, was this parent entangled in a divorce/ custody dispute with the alleged abuser? Was the child pressured to make the disclosure?

(d) Other Evidence. Was there medical/physical evidence consistent with the alleged abuse? Is the child's statement consistent with other statements made by the child and/or other witnesses? Was there material evidence supporting the allegation, and behavioral evidence consistent with abuse (e.g., changes in sleeping/eating patterns, sexual acting out)?

The Validity Checklist is a very important part of SVA, and research must be carried out to evaluate its usefulness. The

present study was designed specifically to examine the reliability and validity of CBCA evaluation. The Validity Checklist was not entirely relevant to the type of contrived eyewitness event and recall task used in this study. Therefore, the Validity Checklist was not applied to the statements obtained.

4.8 Using CBCA

CBCA was developed as a qualitative evaluation procedure. The criteria can be judged to be present or absent, or can be rated on a 4-point scale in terms of the extent to which they are fulfilled (i.e., 0 = not present to 3 = strongly present, Stellar & Koehnken, 1990). In research investigations of CBCA, the criteria are rated numerically. However, in the field, criteria are not numerically rated, rather they are simply judged as present or absent on the basis of the evaluator's impressions of the statement and its contents (Yuille, 1990d).

Yuille (1990d) reported that in the forensic context, expert CBCA-based opinions about the credibility of children's reports have not been based on the application of rigid decision rules with respect to the number of CBCA criteria met. Rather, experts have loosely followed guidelines set out by the developers of CBCA. At least two different sets of guidelines have been proposed. Yuille (1990d) recommended that in order to categorize a statement as likely credible, the first five CBCA criteria plus any other two of the remaining content criteria be fulfilled. Thus, Yuille's guidelines consider Criteria 1 through 5 to be essential in a true account, whereas Criteria 6 through 19 are considered one-sided in

their application. That is, for Criteria 6 to 19, the presence of an individual criterion may be seen as enhancing the validity of the statement, but its absence does not necessarily detract from the statement's credibility. Alternatively, guidelines set out by both Raskin and Stellar suggest that a statement is to be judged as likely credible if the first three content criteria plus any four additional criteria are met (Yuille, 1990d). The difference in decision guidelines (i.e., Yuille versus Raskin and Stellar) reflects the troubling fact that no one really knows exactly what factors, or how many such factors, are critical for determining a statement's credibility. It is likely that some content criteria are more, or less, significant than others in distinguishing credible from noncredible statements. However, research to determine appropriate cut-off scores and/or empirical weightings of the content criteria for predicting credibility has not been conducted.

In the present study, Yuille's guidelines (i.e., first five criteria--minus Criterion 4, Contextual Embedding, which was not relevant to the experimental scenario--plus any other two criteria) were used to make decisions regarding credibility. However, unlike the loose application of these guidelines in the field, Yuille's guidelines were treated as a clear-cut decision rule in the present experimental investigation. Of note, the two criteria added on an exploratory basis (i.e., Criterion 20, Other's Action, and Criterion 21, Own Action) were not included as criteria entering into the

credibility decisions[1].

CBCA is presently being used to evaluate children's testimony in forensic practice. The method clearly reflects a great deal of expert knowledge based on practical experience. However, there has been a paucity of systematic scientific research to test its reliability and validity as a credibility assessment tool. Evidence supporting the usefulness of statement analysis consists mainly of unsystematically gathered case reports (e.g., Arntzen, 1982, 1983, cited in Stellar & Koehnken, 1990; Undeutsch, 1982). At present, it appears that the clear description of, and reasonably well explicated distinctions between, the content criteria should make it possible to conduct systematic empirical research to examine the reliability and validity of CBCA. In the remaining sections of chapter 4, the field and experimental research regarding this statement analysis approach are reviewed and conclusions of the studies are discussed.

4.9 CBCA: Research Investigations

4.10 Field Studies. The only field validation study of CBCA to date was conducted by Esplin, Boychuk, and Raskin (1988, cited in Raskin & Esplin, 1991, and Stellar, 1989). Forty children, aged 3-1/2 to 17 years, referred to a psychologist because of alleged sexual abuse were interviewed. For 20 of the children, allegations of abuse were confirmed by medical evidence, deceptive polygraph outcomes, and/or a confession by the alleged abuser. For the other

[1] In order to ensure that inaccurate credibility decisions were not being made because a nonoptimal decision rule was being used, analyses were carried out in which classification accuracy achieved using Yuille's guidelines was compared with that achieved using continuous CBCA scores (i.e., summed scores on all CBCA content criteria; see Section 7.2).

20, abuse was unconfirmed (i.e., lack of medical or other corroborating evidence, non-deceptive polygraph outcomes, alleged abuser denied allegations, clinical judgment by the psychologist that abuse was unlikely to have occurred, and judicial dismissal). Interviews were transcribed, then evaluated according to CBCA by a trained rater who was blind to whether the abuse was confirmed or unconfirmed. Each of the 19 content criteria was scored on a three-point rating scale (0 = absent, 1 = present, 2 = strongly present), for a total possible CBCA score of 38.

Results indicated that the two groups were clearly differentiated by CBCA evaluation. The mean score for children in the confirmed group was 24.8, while children in the unconfirmed group scored an average of 3.6 points. Some of the individual criteria were found to strongly differentiate group membership. In fact, some of the criteria met by a relatively high percentage of statements by children in the confirmed group were completely absent in statements of the unconfirmed group (particularly Reproduction of Conversation, Unexpected Complication, Unusual Details, Related External Associations, Attributions of Perpetrator's Mental State).

Esplin et al.'s field study does seem to bode well for the validity of CBCA. However, this study was not without flaws. The 40 cases used in the investigation were all obtained from cases referred to two psychologists (two of the authors), therefore the representativeness of the sample for the general population of sexual abuse cases is questionable. Further, transcripts were evaluated according to CBCA by only one individual. Thus, this investigation did not address the reliability of CBCA across raters.

It is notable that although the children varied considerably in age, the investigators did not address the relationship between age of the child and the complexity of the event. As Stellar (1989) pointed out, this age factor may have had an unrecognized effect on the strength and quantity of criteria fulfilled, and on the overall quality of the statements.

The most serious problem of this study, as well as of most any attempt to conduct a field validation study of CBCA, was that of determining the criterion by which the validity of the procedure was assessed. In actual criminal cases, particularly those involving sexual abuse, there is most often no "simple, objective, independent, and reliable criterion", or ground-truth criterion, by which to determine exactly what did or did not occur (Unduetsch, 1984, p. 64). In Esplin et al.'s field study, it is possible that some cases included in the group of unconfirmed allegations were actually credible allegations misclassified on the basis of a lack of ground-truth criteria. Further, cases confidently classified as confirmed may in reality have been actual cases of abuse carried out, not by the alleged perpetrator but, by someone the child had chosen to protect.

In another field study, Anson (1991) had trained raters apply CBCA to a sample of 23 videotaped interviews of confirmed sexual abuse cases (ranging in age from 4 to 12 years). The mean CBCA score for the rated videotapes (i.e., sum of scores on each of the 19 criteria, each rated on a scale of 0 to 2) was 10.4, with a range of 1.7 to 17.5. This mean CBCA score is strikingly lower than the score of 24.8 obtained for the confirmed sexual abuse group in

Esplin et al.'s (1988) study. Anson suggested a number of possible explanations for this difference. These included differences across studies in the ages of subject samples, number of prior interviews, nature of the abuse and degree of injury suffered, rates of free-narratives provided by children, rating of videotaped interviews versus transcripts, degree of control over interview style, and number of CBCA raters used. Whereas Esplin et al. had one CBCA rater with "extensive training in CBCA" (Raskin and Esplin, 1991, p. 161), CBCA raters in Anson's study had taken a university course on Statement Validity Analysis interview and assessment procedures and a weekend workshop on the application of CBCA. Unfortunately, although Anson intended to include a group of unconfirmed sexual abuse cases in his study, he was unable to do so. Out of the total number of alleged cases classified as probably/definitely false, only two contained an allegation of abuse during the videotaped interview. Thus, he was unable to carry out a test of the validity of CBCA for discriminating between credible and non-credible allegations of abuse.

Yuille (in press) is presently completing a field research project designed to evaluate the validity of SVA for assessing the credibility of children's testimony in abuse cases. Three Vancouver-area communities participated in this project. Initially, professionals at two Royal Canadian Mounted Police detachments and social service districts were trained in the use of the Step-Wise Interview and SVA. Training consisted of a 4 day workshop covering theory and practical applications of both the interview and statement analysis procedures. Workshops were presented to groups

of approximately 30 professionals. Following the completion of training, reported cases of child sexual abuse in which formal investigative interviews were conducted and taped (audio or video), and for which parental consents were obtained, were provided to the field project. In total, 233 interviews were transcribed and evaluated according to SVA. A third community served as a control site. The professionals in this district did not receive training for the first 6 months of the project, but did provide to the project their reports of child sexual abuse and the accompanying taped interviews. At the conclusion of the 6 month period, professionals in this district were trained, thus allowing for a between groups comparison of the adequacy of interviewing methods and credibility assessment decisions (i.e., between the trained and untrained sites), and a pre- to post-training within group comparison of the performance of professionals at this third site.

The results of this study are not yet available. However, Yuille (in press) reported very promising preliminary results with regard to the improved quality of investigative interviewing with the use of the Step-Wise Interview. With regard to SVA, he reported only that the majority of professionals trained had difficulty with the application of the statement analysis procedure. As he pointed out, this finding supports Undeutsch's assertion that statement analysis is a difficult procedure requiring special training and skill.

4.11 Experimental Studies. In reaction to the problem of establishing the ground truth criterion, as well as to other

methodological shortcomings inherent in field studies (e.g., lack of randomization and variable control), Stellar and Koehnken (1990) recommended that experimental methods be used for initial investigations of CBCA. Stellar (1989) proposed that experimental studies can provide useful information if the target event directly involves the witness, has a predominantly negative emotional tone, and involves an extensive loss of control over the situation. Raskin and Esplin (1991) also expressed support for experimental investigations of CBCA, but cautioned that the target event need not have a negative tone, as the tone is not necessarily negative in cases of sexual abuse. They suggested that the event should incorporate novel aspects, in order to make it unlikely that a child could fabricate a credible sounding report on the basis of prior experience with events similar to that of the experimental event. Further, Raskin and Esplin (1991) strongly stated that in order to consider application of research findings to the sexual abuse context, it is necessary to obtain statements from children who have been motivated to make misrepresentations that they believe will be accepted by an adult.

A number of other authors (e.g., Arntzen, 1983; Trankell, 1971, cited in Stellar & Koehnken, 1990) have severely criticized the use of experimentation to investigate the usefulness of CBCA. Arntzen (1983, cited in Stellar & Koehnken, 1990) asserted that experimental investigations are of no worth for the evaluation of statement analysis because they are artificial and lack the kind of significant personal and emotional involvement typically found in sexual abuse cases.

It is true that, by their very design, experimental investigations are contrived and aim to be gentle in emotional impact. Two of the three criteria suggested by Stellar (1989) as necessary for experimental investigations (i.e., negative tone, extensive loss of control) can be ethically and practically impossible to implement in experimental investigations with children. Although the absence of these features may limit generalizability of findings from 'lab' to field, it does not render experimental studies inappropriate for investigating the validity of statement analysis procedures.

CBCA was developed as a means of explicating the qualitative criteria distinguishing credible from noncredible witness statements. If the basic assumption (i.e., the Undeutsch Hypothesis) underlying this method is justified, it should hold for topics outside of the sexual abuse arena (Stellar, 1989). Thus, although some of the 19 qualitative criteria (e.g., Criterion 10, Accurately Reported Details Not Understood; Criterion 18, Pardoning the Perpetrator) may not be applicable to testimony derived from an experimental investigation, it should be possible to apply the overall statement analysis procedure to experimental events that are personally involving for the child witness. In fact, with features of intense emotional involvement and negative tone absent in the staged event, experimental investigations may provide a more conservative test of CBCA's ability to discriminate true and false statements than is possible in the field. From a methodological perspective, such conservatism in the early stages of validating a procedure is to be lauded. When statement analysis has been

thoroughly investigated in highly controlled experimental investigations, generalizability of findings to the forensic context can be assessed through field studies (Stellar & Koehnken, 1990).

At present, there are few experimental investigations of the validity of CBCA. Koehnken and Wegener (1982, cited in Stellar, 1989) analyzed the statements of adolescent girls (aged 16 to 17 years) with regard to three content criteria: number of details, spontaneous reproduction, and coherence over repeated questioning. Half of the subjects were shown a 10 minute film depicting a family argument, the other half were given a verbal description of the contents of the film. Subjects were interviewed, and transcripts of the interviews were rated by trained raters blind to the experimental hypotheses and subjects' experimental conditions.

Subjects who saw the film produced significantly more detail than did subjects in the fantasy group. Contrary to expectation, spontaneous reproduction was found more often in the fantasy group, and no differences were found between groups in the consistency of the content of their reports over repeated questioning. The results of this investigation, though interesting, must be viewed as preliminary, for only three content criteria were assessed and the witnessed event (by virtue of being presented on film) was not personally involving.

Stellar, Wellershaus, and Wolf (1988, cited in Stellar, 1989) conducted a simulation study in which children in Grades 1 and 4 were instructed to tell two stories, one based on a personally experienced event and one that they had invented. The story themes were to be selected from a number of topics thought by the

experimenters to characterize key variables (i.e., direct involvement, loss of control, negative emotional tone) of sexual abuse. These topics included situations in which the child received medical treatment, or non-medical topics such as being beat up, being attacked by a dog, etc. Parents served as 'objective criteria of reality' by providing information about the actual events experienced by their children.

Stellar et al. found that CBCA adequately distinguished between true and false stories on medical topics, but not for non-medical topics. When a follow-up assessment was done on only the stories with medical themes, 11 out of the 17 criteria differentiated significantly between true and false stories. (Note: 2 of the 19 criteria were not applied because they were not relevant to the nature of the events reported). Criterion 2, Unstructured Production, Criterion 13, Attribution of Perpetrator's Mental State, and all criteria of the fourth content category (i.e., Motivation-Related Contents), failed to differ between the true and invented stories. Further, the researchers demonstrated that raters trained for only 90 minutes in the use of CBCA, made significantly more correct credibility classifications than did untrained raters who relied on intuitive judgments. They reported that for true and false reports combined, CBCA correctly classified 71.9%, and untrained raters 60%, of the reports. For true reports, CBCA accurately classified 77.7% relative to the 68% correct classifications by untrained raters. For false reports, CBCA correctly classified 62.3% and untrained raters 47% of the reports. Stellar (1989) reported these findings as "proof that use of CBCA

... enhances the correct credibility classifications of children's statements about topics which show some features similar to the sexual abuse contact" (p. 149). Further, he recommended future efforts to clarify the nature and characteristics of events for which the Undeutsch Hypothesis is valid.

Briefly, it should be noted that the apparent inability of CBCA to accurately distinguish between true and false accounts of non-medical topics may have had more to do with shortcomings in the methodology of this experiment than with the fact that these events are substantially different than episodes of sexual abuse. Specifically, the very brief period of CBCA training given to raters could well have resulted in inadequately trained raters who, although able to recognize obvious discriminating factors, were not sufficiently skilled in the use of the procedure to make more subtle discriminations. It is possible that the accounts of medical experiences included more of the obvious discriminating features than did reports of the non-medical experiences. Another possible explanation for the apparent failing of CBCA in distinguishing between true and false accounts of non-medical experiences is that the parents, serving as informants regarding the actual experiences of their children, may have simply lacked knowledge of their children's non-medical experiences.

Landry and Brigham (in press) conducted a simulation study similar to that of Stellar et al.'s (1988) study, but the statements analyzed were made by adults. University students were videotaped giving 1 to 2 minute descriptions of two personal incidents that were traumatic, emotionally involving, and during which they felt a

loss of control. For each student, one of the two incidents was to be a true personal experience and the other (the topic of which was assigned by the experimenter 2 days prior to videotaping) was to be invented. Twelve videotaped statements (six true experience, six invented) were selected for credibility evaluation on the basis of moderately high ratings for degree of trauma, emotional involvement, and loss of control. Credibility was assessed for each of the 12 statements by groups of undergraduates who either received a 45 minute training session in the application of CBCA or received no CBCA training, and who evaluated either videotaped or transcribed versions of the statements.

Landry and Brigham reported that 10 of the 14 CBCA criteria assessed were present significantly more often in the truthful statements. Two criteria (Criterion 1, Logical Structure; Criterion 13, Other's Mental State) were more often met in false statements. Two additional criteria (i.e., Criterion 7, Unexpected Complications; Criterion 17, Self-deprecation) did not differ across conditions. They found a significant difference in the accuracy rate of credibility decisions between CBCA-trained versus untrained raters (55.3% vs. 46.9%). As well, videotaped presentation resulted in a significantly higher rate of accurate classifications than judgments made from transcribed statements (50.2% vs. 42.5%). CBCA-trained raters who viewed videotaped statements had the highest accuracy rate of all groups (58.1%, or 52% when videotapes for which raters were unable to decide on credibility status were included in the analysis).

Landry and Brigham interpreted their results, particularly the finding that raters in the training-videotape condition performed significantly better than chance, as support for the validity of CBCA. Such a conclusion is problematic in light of the fact that, even in this condition, raters inaccurately classified the credibility of transcripts 41.9% of the time. While demonstrating that their raters performed better than chance may be statistically significant, the clinical significance of this finding is questionable, and certainly does not provide strong support for the validity of CBCA.

As a validation study of CBCA, this investigation had two serious limitations. CBCA raters were trained in the use of CBCA in only one, extremely brief, session. In spite of this limited exposure to CBCA, raters were not permitted to refer to any notes or handouts on the specific criteria as they judged the 12 statements. Their resulting credibility decisions, therefore, were unlikely to reflect adequate application of the procedure. Thus, findings of this study regarding the validity of CBCA for distinguishing between true and false statements must be viewed with caution. As well, CBCA's difficulty in distinguishing between true and false statements by adults cannot be assumed to reflect its potential usefulness with statements by children. Adults, with their more highly developed cognitive/memorial abilities and more advanced knowledge of what constitutes a believable statement, may simply be better able than children to fool CBCA by incorporating features they associate with credibility into their statements. In fact, Landry and Brigham's finding that Criterion 1, Logical Structure,

was met more frequently in the false statements may support this possibility.

Yuille (1988b) conducted a simulation study similar to those of Stellar et al. (1988) and Landry et al. (in press). Children, 6 to 9 years of age, were given 2 days notice that they would be asked to tell two stories, one of which was to recount a true experience and the other of which was to be fictional (but plausible). Two days later, children were interviewed using the Step-Wise Interview by interviewers unaware of which stories were true and false. Two blind evaluators (i.e., undergraduates trained in the application of CBCA in an intensive weekend workshop) assessed the transcript of each story according to CBCA. The two evaluators agreed on 96% of their classifications. Overall, there was a 90.9% level of correct classifications for true stories, and 70.4% correct classifications for false stories.

This rate of correct classifications was markedly better than the rate (55.3% overall) reported by Landry and Brigham. Nevertheless, Yuille's CBCA evaluators' Type I error rate (i.e., labeling credible statements as not credible) of 9.1% and Type II error rate (i.e., labeling false statements as credible) of 29.6% raise some doubt as to the adequacy of CBCA classifications. However, a number of factors would suggest that this study be considered a conservative test of the accuracy of CBCA classifications.

First, although CBCA raters in this study received more hours of training in the application of CBCA than did raters in the previous two experimental studies reviewed, it is still possible

that Yuille's raters were not sufficiently familiar with the procedure to apply it optimally. Second, the CBCA evaluation did not investigate the usefulness of all 19 content criteria in assessing credibility, as many of the criteria could not be applied to the children's statements because they were not relevant to the type of innocuous events reported. Third, similar to the major problem in field investigations, there was no available ground-truth criterion by which to determine that the children's reports of experienced events were true to the facts of the incident, or that the reported incidents were cognitively and emotionally involving for the child.

Finally, post-interview questioning revealed that many of the children's fictional stories were based on truly experienced events. Thus, it is possible that the level of correct classifications for false stories would have been higher if the children had been able to give fictional accounts. This unexpected finding may have implications for the area of children's false allegations of abuse. While it is clear that children can and do falsely allege abuse, the impetus for their allegations is not always clear. Yuille's finding that children had difficulty inventing a 'memory' for a nonexperienced event raises the possibility that children who succeed in giving plausible false accounts of abuse could be basing their reports on actual experiences that have been in some way distorted. Alternatively, they may be relying heavily on the coaching of adults, with the cognitive capacity and knowledge base to invent believable fictional accounts.

Yuille (1991) conducted another pilot investigation of the validity of CBCA for distinguishing between children's reports of true and fictional experience. In this study, classes of children (Grades 2 to 5) either witnessed a staged event, which took place in front of the whole class, or heard a narrated account of the same event. The event itself was innocuous, involving a heat inspector's check of the temperature in different areas of the classroom. One day later, children were interviewed for their recall of the event. Children in the narrated condition had been forewarned that they were to present their recall for the event as though they had really experienced it. Taped interviews were transcribed and the transcripts were scored according to CBCA criteria. Again, Yuille's CBCA evaluators were undergraduates trained in the application of CBCA in a 2 day workshop.

Although the results of this investigation have not been fully analyzed, Yuille (1991) reported that the total amount and accuracy of recalled information did not differ between experimental conditions. Children who were simply told about the event remembered as much, and as accurately, as the children who witnessed the staged event. As might be expected, Grade 4 and 5 students remembered more detail that did Grade 2 and 3 students; however, this difference was not statistically significant. Based on preliminary analyses of the CBCA results, Yuille reported that children's reports of live versus narrated events were 100% correctly classified by the statement analysis procedure.

These preliminary findings make a strong case for the value of continued experimental investigation of CBCA. The methodology used

in Yuille's (1991) pilot study constituted an important improvement over the methodology used in his previous (1988b) pilot investigation of CBCA's classification accuracy. By providing children with the fictional account to be recalled, the 1991 pilot study bypassed the problem of children's potential difficulty with inventing a fictional event. In addition, the methodological change brought the experimental methodology a step closer to simulating situations in which children's false testimony is based on adult coaching.

There were a number of methodological shortcomings in Yuille's (1991) pilot study that should be dealt with in future research. First, the three conditions proposed by Stellar (1989) as critical for generalizability of findings from laboratory to field were not adequately met in this study. Because the staged event was innocuous, and was presented in front of a large number of children, there was little chance of children becoming personally (i.e., emotionally or behaviourally) involved in the event. There was no loss of control, and certainly no negative emotional tone. Further, the fact that large numbers of children were exposed to the event at one time made it impossible to observe and judge whether individual children were attending to the event, or even whether they were in a position allowing them to adequately witness the event.

Second, although Yuille's (1991) pilot study provided an initial look at the effectiveness of coaching for enabling children to later recount a fictional event as though it was actually experienced, the effectiveness of the deception instructions to children in the narrated condition was not examined. Prior to

hearing about the event, these children were told that they were to present their later recall for the narrated event as if they had actually witnessed it. As well, they were reminded not to make comments that would give away the fact that they had not truly witnessed the event. Any such giveaways were edited out of the transcripts of children's statements before CBCA evaluation. At a theoretical level, this editing allowed the investigator to carry on with the assessment of other qualitative differences between credible and noncredible reports. But, at a practical level (particularly when generalizability to the forensic context is considered), there is really no reason to apply the statement analysis procedure to accounts with glaring indications of being false. Thus, it is important to ensure that the instructions to children in coached conditions are maximally effective for enabling them to present recall of coached events in a way that, at least superficially, does not betray their attempt to mislead the interviewer.

Third, Yuille did not attempt to quantify and/or vary the levels of coaching used. Further research investigating the effects of varying levels of coaching (for example, the provision of minimum versus comprehensive details--including qualitative features designed to meet CBCA criteria-- about the to-be-recalled event, single versus multiple presentations of coaching, no practice versus practice sessions) on determinations of credibility would be of value. Finally, although the reported results of Yuille's (1991) pilot study did not call into question the training or performance of the CBCA evaluators used, future research on CBCA must carefully

consider the amount of training and experience/proficiency of evaluators applying the procedure in order to make claims about the validity of CBCA as a system for judging the credibility of eyewitness statements.

4.12 Conclusions and Future Directions for Investigations of the Validity of CBCA. CBCA and its predecessors have been used for decades to assess the credibility of children's statements in cases of alleged sexual abuse. There have been, though, only a handful of experimental and field investigations of the validity of CBCA. These studies hint that CBCA is a useful system for assessing the credibility of children's testimony. However, its validity has not yet been adequately empirically tested.

The only published field validation study to date suffers from serious methodological limitations, including a selected sample, lack of attention to the influence of age on the quantitative and qualitative characteristics of children's statements, and the problem of determining an objective criterion of reality. Since applications of CBCA are primarily forensic, continued field research is critical to the eventual validation of this credibility assessment procedure.

Methodologically sound field investigations are expected to add to our knowledge of the scientific basis for SVA (including CBCA). As previously discussed, though, the confidence with which conclusions about the validity of CBCA can be drawn on the basis of such research is tempered by the problem of determining the ground-truth criteria. It is often impossible to determine what truly

happened, or did not happen, in cases of alleged sexual abuse. Thus, the criteria by which the accuracy of CBCA is judged may be incorrect, resulting in erroneous conclusion regarding the validity of CBCA.

Experimental investigations offer another avenue for exploring the validity of CBCA. The high degree of control possible in experimental studies allows the experimenter to ensure that the ground-truth criterion is very clear. In turn, this absolute knowledge of the ground-truth criterion makes it possible to verify the accuracy of CBCA classification decisions.

Despite this positive characterization of experimental investigations, the research to date has not succeeded in providing a rigorous test of the validity of CBCA. This lack of success can be largely attributed to methodological shortcomings in the studies. Investigators have had difficulty creating to-be-witnessed events that were involving for children, and that were not so innocuous as to make CBCA evaluation of the children's eyewitness reports meaningless. In addition, results have been left open to interpretation because no attempts were made to assess whether the children were attending to and could adequately witness the stimulus event. In studies assessing discriminations between true and false reports of past experiences, there have been inadequacies in the ground-truth criteria used to determine the credibility of the children's reports, and problems related to the children's inability to present reports that were not based on personal experiences. In the investigation of children's accounts of an experienced versus a coached event, difficulties were discovered in ensuring that

children in the coached condition understood, and were motivated to comply with, instructions encouraging them to deliver a false report to an unsuspecting interviewer. Further, CBCA evaluation was, in some studies, carried out by individuals insufficiently trained in the procedure, and/or by only one evaluator (with no attention to the importance of demonstrating inter-rater reliability).

This enumeration of flaws in the present body of experimental research on the motivational aspects of the credibility of children's testimony need not deter future experimental investigations on this topic. To the contrary, previous research efforts have set the stage for more sophisticated and methodologically sound experimental research. Since the negative tone, loss of control, and degree of emotional involvement characteristic of real-world forensically relevant events cannot, and decidedly should not, be simulated in experimental investigations with children, it will be important to be cautious about generalizing future experimental findings to the forensic context. A productive strategy would be to assess the results obtained in experimental investigations in relation to those obtained in the field. Then, through converging operations, we can gain a more complete understanding of the validity of CBCA for assessing the credibility of children's eyewitness reports.

Chapter 5

THE PRESENT STUDY, HYPOTHESES, AND METHOD

5.1 The Present Study

To date, many European court decisions have been influenced by credibility assessments based on the assumption that the Undeutsch Hypothesis is correct, and that the qualitative characteristics identified by Undeutsch (and explicated in CBCA) enable accurate discriminations of credible and noncredible reports to be made. However, the validity of the Undeutsch Hypothesis and CBCA have not been adequately tested.

The goal of my thesis was to experimentally test the Undeutsch Hypothesis which states that accounts based on memory for experienced events will be distinguishably different, qualitatively and quantitatively, from accounts based on fantasy/coaching. In order to test this hypothesis, eyewitness reports were obtained from children of two different grades (Grades 2 and 4) in three experimental conditions. In one condition, Live Event (LE), children individually witnessed and participated in a staged event. The event was complex, involved the children directly in interactions/conversation with the confederate actor posing as a repairman, and included a number of features considered by CBCA to be relevant to credibility (e.g., an unexpected interruption, unusual detail, obvious emotional reaction displayed on the face of the repairman). Two other groups of children did not witness the event, but were individually coached about it. In one of the two

coached conditions, the Heavily Coached (HC) condition, the children saw a picture of the repairman and received a detailed oral account of the event, including features which--if reported--would be assigned credibility enhancing significance by CBCA. In the other, Lightly Coached (LC) condition, children received a more skeletal account of the event, covering basic persons, objects, and actions involved in the live event, but leaving to the children the task of filling in the details that would make their reports believable. Before hearing about the event, children in the coached conditions were encouraged to pretend (as they listened to the coaching) that the event had happened to them, and to later attempt to fool the interviewer into believing that their reports were based on personal experience. Recall for the event was obtained through individual interviews which took place immediately after the event presentation. Verbatim transcripts of the interviews allowed quantitative and qualitative characteristics of the children's statements to be assessed.

Two age groups of children (i.e., Grades 4 and 2) were selected in order to assess the robustness across different ages of any emergent differences in the quantitative and qualitative characteristics of children's reports based on experience versus coaching. It would have been ideal to test children from all primary school grades, but practical considerations determined that only two grades be selected. I attempted to select grades of children sufficiently different in age to allow potential age differences to emerge. At the lower end, my experience piloting the procedures with one class of Grade 2 students led me to be

pessimistic about the capacity of children below Grade 2 to understand the task demands, to present their coached reports without blatant giveaways, and/or to present enough detail in their account to make CBCA evaluation possible. Thus, children in Grade 2 were selected as the young sample.

At the higher end, Grade 4 students were selected because I was concerned that past age 10 to 11, children would too readily recognize the staged nature of the LE and would therefore respond with amusement and half-hearted cooperation, rather than with earnest attention and full cooperation. Results of past research on children's eyewitnessing abilities (see Ceci, Toglia, & Ross, 1987) suggested that the difference between the two age groups of children in the present study would be large enough to allow age differences to emerge.

The two types of coaching (i.e., HC and LC) were included for different reasons. The HC coaching was designed to provide rich detail of the LE, specifically including a number of features suggested by Undeutsch to reflect true eyewitness accounts and which, if reported, would meet CBCA content criteria (i.e., would be assigned credibility enhancing significance by CBCA). HC was not intended to represent the kind of coaching provided to children in the real world. Adults who coach children to provide false testimony in forensic cases are unlikely to be familiar with CBCA criteria, therefore cannot be expected to deliberately train children to incorporate features meeting the qualitative criteria of CBCA. Thus, to judge the validity of CBCA on the basis of its efficiency at accurately classifying LE versus HC transcripts would

not be a fair test of the procedure. Instead, the HC condition was intended to provide a critical test of an important boundary condition of CBCA. That is, can children be coached to provide information meeting CBCA criteria? Further, is the CBCA evaluation procedure sophisticated enough to discriminate credible statements from noncredible reports that are the product of coaching tailored to meet CBCA criteria?

A more reasonable test of the Undeutsch Hypothesis, and the efficiency of CBCA for distinguishing credible and noncredible reports, would be the success of CBCA evaluation in classifying LE versus LC transcripts. The coaching delivered in the LC condition was NOT designed to incorporate features meeting the qualitative criteria of CBCA. Thus, LC likely more closely approximates the level of coaching provided to children in the forensic context, leaving to the children the challenge of including features that would make their reports believable.

Efforts were made to overcome a number of methodological shortcomings identified through the work of previous investigators (and detailed in sections 4.11 and 4.12). Specifically, the ground-truth criterion (i.e., in this case, experimental condition) by which the accuracy of CBCA determinations of credibility were established was well known by the investigator. Presentation of the to-be-remembered event was highly standardized within conditions. The individual nature of the staged event, and the fact that participation was requested of the child witnessing the event, provided some assurance that the child was attending to, and involved in, the event. Similarly, the individual coaching sessions

ensured, at minimum, that obvious signs of a child's inattention could be recorded.

No attempt was made to directly map features of sexual abuse onto the event scenario. An effort was made, though, to meet, or at least approximate, the standards set out by Stellar (1989) and Raskin and Esplin (1991) for experimental investigations of CBCA. The event was intended to be directly involving, have a component of lost control, have a somewhat negative tone, and involve novel aspects unlikely to be thought of in a fabrication. Pilot testing enabled the development of an event that was involving, and incorporated some negative tone and lost control, but was not upsetting for the children. As well, pilot testing led to the development of instructions that were effective in motivating children in the coached conditions to attempt to make their false reports believable to the interviewer, and in enabling the children to present their recall without tell-tale signs of their experimental condition.

Specific purposes of the present study

Within the overarching aim of testing the validity of the Undeutsch Hypothesis, this investigation served a number of specific purposes. The first and primary purpose of the study was to provide a direct experimental test of the validity of CBCA for discriminating between credible and noncredible eyewitness reports. The accuracy of decisions made using CBCA were compared across the three experimental conditions.

Second, this study permitted a comparison of the number of criteria met, and the degree to which these criteria were fulfilled, across conditions and grades. Third, it served as an initial investigation of the relative contribution of the different categories of CBCA content criteria to making discriminations between testimony based on personal experience versus coaching. Fourth, an assessment of which individual content criteria differed significantly across conditions was included.

Fifth, this thesis involved a preliminary investigation of whether credibility decisions made using CBCA differ in accuracy from credibility decisions made by individuals untrained in CBCA. The accuracy of CBCA classification decisions was compared, for a subset of Grade 4 LE and HC transcripts, to the accuracy of classification decisions made by untrained evaluators (i.e., adults who have not had experience making professional judgments of credibility, and who are unfamiliar with CBCA)[2]. The comparison of evaluations by CBCA and untrained evaluators was designed specifically to test the possibilities that (a) the features discriminating true and false accounts are so obvious as to make CBCA evaluation unnecessary, (b) although not providing perfect discrimination, CBCA's hit rate is superior to that of untrained

2 At the time that transcripts were to be randomly selected for distribution to untrained evaluators, the CBCA evaluation results were not yet available. I was concerned that the obvious differences in the length of the LE/HC transcripts from the LC transcripts would lead untrained evaluators to focus exclusively on the differential length of transcripts in making their judgments. Thus, I decided to use LE and HC transcripts to the exclusion of LC transcripts for this comparison. Further, I was concerned about the onerousness of the task for the individuals who volunteered to serve as inexperienced evaluators. For this reason, I decided to use only Grade 5 transcripts rather than doubling the inexperienced evaluators' work by including a sample of Grade 2 transcripts.

evaluators, and (c) untrained evaluators are able to make better discriminations than is possible when one is rigidly adhering to the systematized CBCA procedure.

Sixth, this study permitted an evaluation of two cognitive aspects of credibility (i.e., amount of information, accuracy of information) in relation to children's recall of live and coached events. Although these quantitative characteristics of memory have been the topic of numerous investigations, there has not yet been a direct comparison of the amount and accuracy of information in children's reports based on experience versus coaching. Such a comparison is of theoretical relevance to questions regarding the effect of live versus coached presentation of events on children's recall. Further, this test provided a way of gaining practical information regarding whether it may be possible to enhance the accuracy of credibility assessments by attending specifically to the amount and accuracy of detail provided. This evaluation is clearly more relevant to experimental research than to the assessment of children's statements in the forensic context. In real-life cases requiring eyewitness testimony by children, the individual nature of the events in question, the common lack of corroborating evidence and/or witnesses makes it difficult, if not impossible, to evaluate the amount and accuracy of detail provided.

Finally, this study provided an opportunity to test the generalizability of the experimental work of Johnson and her colleagues (e.g., Johnson, 1988) and Schooler and his colleagues (Schooler et al., 1986, 1988) to children. As previously discussed, these researchers identified a number of features which

distinguished between adults' reported memories for real versus suggested/imagined events. Lindsey and Johnson (1987) hinted that these differences may be identifiable in children's reported memories for real versus imagined events as well. However, there has been no experimental test of applications to children's memories.

The present study includes an exploratory evaluation of these features in the reports of children. However, since this evaluation was included on an exploratory basis and the coding involved proved to be labour intensive, it was decided that the coding would be applied to only a subset of transcripts. The HC coaching script is most similar, in degree of detail provided, to the script used by Johnson et al. (1988) in having subjects imagine events to later be recalled. Thus, Grade 4 children's reports based on LE and HC presentations were selected for evaluation according to a number of the discriminating characteristics identified by these researchers (i.e., number of sensory details, references to cognitive operations, self-references, contextual details, and verbal hedges). This comparison was expected to provide theoretically interesting information about the applicability of Reality Monitoring findings to children. At a practical level, this evaluation was intended to provide information about statement characteristics that could be of value for enhancing the accuracy of discriminations between true and false reports by children.

5.2 Hypotheses/Research Questions

Hypothesis 1: CBCA Classification Accuracy

(a) LE versus LC. On the basis of the Undeutsch hypothesis, it was predicted that CBCA evaluation would result in highly accurate discriminations between reports by children in the LE and LC conditions across both grades. More specifically, the transcripts of statements by children in the LE condition were expected to be classified as credible, and the statements by children in the LC condition were expected to be classified as noncredible.

(b) LE versus HC. No specific predictions were made. Children in the HC condition were coached in a number of the very details assigned credibility enhancing value by CBCA. Good discrimination between LE and HC transcripts would be expected if (a) coaching fails to enable children in the HC condition to later report the qualitative features meeting CBCA criteria, or (b) CBCA evaluation is sophisticated enough to recognize the differences between presentation of these features based on true experience versus coaching. Alternatively, poor discrimination between LE and HC transcripts (with HC transcripts judged as credible) would be expected if HC condition children can successfully incorporate qualitative features meeting CBCA criteria into their reports.

Hypothesis 2: Number and Degree of Fulfillment of CBCA Content Criteria Met Across Experimental Conditions

The Undeutsch Hypothesis leads to the prediction that the reports of LE subjects, by virtue of being based on an experienced event, would meet more criteria and would receive higher scores on a number of criteria than would the reports of LC subjects. Again, the outcome with regard to HC subjects is questionable. It is likely, though, that even if HC transcripts are inaccurately classified as credible by CBCA evaluation, these HC transcripts would not meet as many of the criteria, nor would they fulfill the criteria met to the same extent, as subjects who actually experienced the event. Thus, it was predicted that there would be increments in the number of content criteria met and in the degree of fulfillment of content criteria for both grades across conditions, with statements by children in the LC condition meeting the least number of content criteria, statements by children in the HC condition meeting more, and statements by children in the LE condition meeting the most criteria.

Question 3: Relative Contribution of Categories of CBCA Content Criteria to Discriminations Between Experimental Conditions

No previous experimental or field research has specifically tested the relative contribution of the categories of content criteria to making discriminations between credible and noncredible statements. Discriminant function analyses were carried out to assess the relative contribution of content categories to

discriminations between conditions, but no specific a priori hypotheses were made.

Question 4: Differences in Individual CBCA Content Criteria Across Experimental Conditions

The analyses conducted to determine which of the criteria significantly differed across conditions were exploratory. No specific a priori predictions were made.

Hypothesis 5: CBCA-Trained versus Untrained Evaluators' Classification Decisions

(a). CBCA evaluation. Based on the Undeutsch hypothesis, it was predicted that CBCA evaluation would result in accurate classification (i.e., credible) of LE transcripts, because of the suitability of CBCA evaluation for identifying the qualitative *criteria of reality* in the statements by children who truly experienced the event. Predictions regarding HC transcripts were more difficult to make. If, in fact, children in the HC condition could be coached to include features meeting CBCA qualitative criteria, errors in the classification of HC transcripts (in the direction of classifying them as credible) would be expected. If, on the other hand, children could not be coached to provide those details, or if CBCA is sophisticated enough to detect that the reports were coached, CBCA evaluation would be expected to result in accurate classification of HC transcripts as noncredible.

(b) Untrained evaluators. It was difficult to make predictions about the performance of this group. However, given

that the untrained evaluators had no experience with making credibility decisions, and were not familiar with the qualitative criteria thought to be important in distinguishing credible and noncredible reports, their performance was expected to be poor with regard to both LE and HC transcripts.

Thus, it was predicted that CBCA evaluation would result in more accurate classification of LE transcripts than would inexperienced evaluation. While inexperienced evaluators were not expected to provide accurate classifications for HC transcripts, no predictions were made regarding the success of CBCA in classifying HC transcripts.

Hypothesis 6: Amount and Accuracy of Detail

(a) Amount of Detail. On the basis of previous research findings (Yuille, 1988b, 1991) and considerations related to the sparseness of detail provided to children in the LC condition, it was hypothesized that the amount of detail provided by children in the LE and HC conditions would be comparable and would be greater than the amount of detail provided by children in the LC condition (LE=HC>LC). Further, on the basis of previous findings of an age-related increase in the amount of information provided by children in free recall reports (e.g., Goodman et al., 1987; King & Yuille, 1987; Marin et al., 1979), it was hypothesized that Grade 4 children (across all conditions) would present more detail than would Grade 2 children.

(b) Accuracy of Detail. Based on the findings of Yuille's (1988b, 1991) pilot studies, no differences were expected in the

accuracy of detail between children in the live and coached conditions. Further, prior research on the accuracy of accounts of an event (see Cole & Loftus, 1987) led to the prediction that there would be no differences in the overall accuracy of reported details by Grade 4 and Grade 2 students.

Hypothesis 7: Exploratory Examination of Johnson/Schooler Qualitative Variables

In accordance with the reported qualitative differences in adults' verbal reports of memories for real versus imagined/suggested events (see Schooler et al., 1986, 1988; Johnson, 1988), it was predicted that statements of children in the LE condition would contain more sensory (i.e., visual and nonvisual) and contextual (i.e. spatial) information, less references to cognitive operations, less self-references, and less verbal hedges than statements by children in the HC condition.

5.3 Method

5.4 Subjects

A total of 172 children (78 Grade 4 and 94 Grade 2) participated in the study. Participants were recruited from eight elementary schools in Vancouver and Richmond, B.C. Prior to including children in the study, a parental consent form (see Appendix A) was sent home by the teacher. Only those children who received parental permission to take part in the research, and who themselves agreed to be involved, participated.

Teachers were asked to identify English as a Second Language (ESL) students and children whose performance on verbal tasks fell below grade level. Those judged to be unable to understand or comply with the demands of the experimental task because of inadequate verbal (receptive or expressive) skills were not included in the final subject sample. Four Grade 4 students were dropped from the final sample, three who were below grade level and one who was an ESL student. An additional 26 Grade 2 students were dropped from the final sample, 10 who were below grade level, 11 who were ESL students, 2 whose testimony included blatant indications of coaching (e.g., "she told me to tell you"), and 3 for whom unexpected incidents made their testimony unusable (e.g., a surprise viewing of the 'repairman' through a window while the child was giving his eyewitness report).

The final sample consisted of 142 children, 74 Grade 4 students (35 boys & 39 girls; mean age = 9.94 years, S.D. = .37) and 68 Grade 2 students (32 boys & 36 girls; mean age = 7.93, S.D. =

.31). Subjects were randomly assigned to three experimental conditions (see Table 2).

1. Live Event (LE). This group consisted of 45 children (26 Grade 4, 15 girls, 11 boys; 21 Grade 2, 11 girls, 10 boys) who witnessed and were actively involved in a staged event.

2. Heavily Coached (HC). This group consisted of 49 children (24 Grade 4, 15 girls, 11 boys; 25 Grade 2, 13 girls, 12 boys) who did not experience the event, but who were told in detail about the features of the event that the children in the LE condition experienced.

3. Lightly Coached (LC). This group consisted of 46 children (24 Grade 4, 12 girls, 12 boys; 22 Grade 2, 12 girls, 10 boys) who did not experience the event, but who were given a skeletal description of the basic items, actions, and interactions involved in the event that the children in the LE condition experienced.

Table 2. Experimental Conditions

LE: LIVE EVENT 47 children (26 Grade 4, 21 Grade 2) witnessed and were actively involved in a staged event.
HC: HEAVILY COACHED 49 children (24 Grade 4, 25 Grade 2) were given a highly detailed account of the event in LE; they did not experience the event.
LC: LIGHTLY COACHED 46 children (24 Grade 4, 22 Grade 2) were given an outline of the event in LE; they did not experience the event.

5.5 Apparatus

Children in all three conditions were taken individually from class and led to a room equipped with a desk, and two chairs facing the desk and angled at approximately 40 degrees from each other. There was a black portable tape recorder on the desk and directly in front of the experimenter's chair. As well, an assortment of pieces of LEGO were scattered on the desk directly in front of the chair on which the child was invited to sit. There was a large table lamp positioned on the floor approximately 6 feet in front of the desk.

In the LE condition, the confederate actor posing as a repairman brought a number of props into the room. He carried a tattered rust colored knapsack. In it, and at various points exposed for the child's viewing, were a 60 watt light bulb, a red pink and white striped towel, an unlabeled cassette tape, and a picture of a kitten. The confederate actor wore black jeans, a white T-shirt, blue high-topped running shoes, and a carpenter's tool belt with a hammer, screwdriver and stopwatch hanging from the belt.

In the HC and LC conditions, the coach read from a script. Children in the HC condition were shown a 4x4 inch photograph of the confederate actor.

Post-event interviews took place in private rooms (i.e., where child and interviewer were not distracted by others and could not be heard by others) furnished with two chairs. The interviewer's equipment included a tape recorder, microphone, blank audio tapes, and clipboard with interviewer instructions/script and blank paper.

5.6 Procedure

In designing this study, a pilot study was conducted with a class of Grade 2 students in order to investigate whether (a) Grade 2 students would be capable of understanding, remembering, and reporting the target event, (b) the manipulations (i.e., partaking in 'theft' for children in LE, and deceiving the interviewer for children in the coached conditions) were upsetting to the children, (c) the coaching instructions contained sufficient incentive to motivate the children to put their best effort into fooling the interviewer, and (d) the coached children would be able to present their recall without blatant giveaways of having been coached. See Appendix B for an elaboration of the pilot study and modifications to the procedures used in the present study on the basis of pilot study findings.

In the present study, the principal investigator entered the classroom and introduced the research project. The children were told that the purpose of the project was to learn more about how children remember things. They were encouraged not to report back to others in the class on their experience in the study until all who had agreed to participate had the opportunity to do so.

Each child was then taken from class individually for an average period of 30 minutes to participate in the study. The child was initially taken to the event room and introduced to the female research assistant who played the role of experimenter or coach. For the first 10 minutes, the child was exposed to either the live event or coaching.

In the Live Event (LE) condition, the child was introduced to the female experimenter and was invited to sit beside her at the desk. Together, the experimenter and child began to build a LEGO house. The experimenter interrupted this activity and temporarily excused herself from the room. Shortly after her exit, a repairman (male confederate) entered the room. He interacted briefly with the child, then attended to a 'broken' lamp. The child was involved in a number of the repairman's activities (e.g., holding a small object for him while he checked the lamp, viewing a picture of the repairman's cat, joining him in testing a tape recorder, and finally helping him pack the tape recorder into his backpack). A number of features in this event were created specifically to meet CBCA criteria, for example, a stopwatch alarm going off (unexpected interruption), wrapping of the tape recorder in a distinctive looking towel (unusual detail). Following this series of activities, the repairman informed the child that he was going to fix the tape recorder and left the room with the tape recorder. The female experimenter returned and asked the child what happened to the tape recorder. Upon learning what had happened, she asked for the child's co-operation in helping her to recover the much needed tape recorder. The child was asked to meet with another researcher and report on everything that happened while the experimenter was out of the room, a period spanning approximately 5 minutes (see Appendix C for a complete description of the LE, including the experimenter's script).

In the coached conditions, the child was introduced to, and seated beside, the coach (i.e., same individual as experimenter in

LE). The coach explained that most of the children participating in the study were to witness an event and then tell an interviewer about it. She further explained that this child's job would be to join the coach in the task of playing a practical joke on the interviewer. The child was asked to do so by first listening to a description of the events that actually occurred when the other children were in the room, then do his/her best to tell the interviewer about these events in such a way as to lead her to believe that the events truly took place in the child's presence. When the child understood the task, the coach read the HC or LC script (see Appendix D for verbatim instructions to children in coached conditions, and for HC and LC scripts).

Immediately following the staged event or coaching, the child was taken to one of two interview rooms and was interviewed by a research assistant trained in the Step-Wise Interview procedure and blind to the child's experimental condition. Precautions were taken to ensure that each of the six interviewers saw approximately equal numbers of children from the two grade levels and the three experimental conditions. Interviews were of approximately 20 minutes duration. (See Appendix E for elaboration of interview protocol and debriefing script). These interviews were audio taped for later analysis. Following the interview and debriefing, the child was returned to the classroom.

Immediately following their involvement with each child, the coach/experimenter, confederate actor, and interviewer completed the appropriate post-event/post-interview questionnaire designed to ensure that any unusual circumstances were noted. The post-

interview questionnaire was also intended to provide easily accessible information regarding whether there were 'giveaways' in the child's testimony making obvious their experimental condition, and whether the child appeared to have language difficulties severe enough to render their testimony unusable in this study.

The six research assistants serving as interviewers attended a 2 day Step-Wise Interview workshop taught by Dr. John Yuille. They then audio taped at least two practice interviews. I met with each interviewer on two to three occasions prior to data collection. In these individual sessions, we reviewed their taped interviews and I provided corrective feedback. Interviewers did not begin interviewing subjects until I judged them to be sufficiently skilled in using the Step-Wise Interview procedure. I continued to meet with the interviewers throughout the first 2 weeks of data collection in order to critically review their taped interviews.

Taped interviews were transcribed verbatim. Transcripts were then evaluated according to the following procedures: (1) CBCA, (2) amount and accuracy of detail, and (3) Johnson/Schooler qualitative variables. In addition, in order to assess whether CBCA evaluation resulted in more accurate classifications regarding the credibility of statements than classification decisions made without this procedure, the accuracy of credibility judgments made on the basis of CBCA was compared with the accuracy of credibility judgments made by individuals with no training in evaluating credibility. This procedure is described below in the section dealing with CBCA.

For each of the above procedures, volunteer undergraduate research assistants served as evaluators. None of the evaluators

were involved in the data collection. Further, they were blind to the hypotheses of the study and to group membership of the subjects. The procedures for evaluating the children's statements are more fully described below.

Evaluation of Children's Recall

(1). CBCA. Initially, 10 psychology undergraduate volunteers were given four 2 hour training sessions in the application of the CBCA. These sessions were conducted by Dr. John Yuille, an expert in SVA and the application of CBCA. Although a recent study by Landry and Brigham (in press) suggested that students could be trained in the use of the CBCA in a single brief (i.e., 45 minute) session, this was not found to be the case in the present study. Students trained in these brief workshops had difficulty judging the coherence, spontaneity, and adequacy of amount of detail in the children's reports, and did not reliably identify components of the reports reflecting the various qualitative criteria. Yuille (in press) reported a similar finding from recent field research. According to Yuille, most professionals (even those already experienced in making decisions about the credibility of children's reports) trained in the use of CBCA during a 2 day intensive workshop did not become proficient at applying CBCA to children's statements.

Thus, it was clear that further training was needed to adequately prepare the research assistants in the present study for their CBCA evaluation task. The initial group of 10 trainees was reduced to the 4 whose homework CBCA coding tasks demonstrated some

aptitude for applying the procedure, and whose schedules coincided enough to make frequent training meetings possible. The 6 remaining volunteers were transferred to the other two coding procedures (to be described below).

Three additional 4 hour training meetings were held with the CBCA trainees, each accompanied by homework coding tasks which were reviewed at a following meeting. I developed a manual with exemplars for CBCA Content Criteria 5 through 19 (minus the criteria which obviously did not apply to the experimental event, and with the addition of criteria 20 and 21). Copies of the manual were given to the coders to aid them in learning to apply CBCA. By the end of the third 'second round' training session, coders were judged to be applying CBCA to children's statements with relative accuracy and reliability. Of note, these early training sessions were focused almost exclusively on application of CBCA to the reports of Grade 4 children. Further, Grade 4 transcripts were all coded before coders began the task of applying CBCA to the Grade 2 transcripts.

When the coding task shifted to the Grade 2 transcripts, two weekly coding meetings focused on reviewing developmental considerations (e.g., expected differences in amount of recall, somewhat less organized reports by younger children) in applying CBCA to statements by this younger age group of children, and on practicing the application of CBCA to Grade 2 statements (most importantly, judging spontaneity and sufficiency of details). When I was satisfied that the CBCA raters were judging Grade 2

transcripts with reasonable competence, coding of these transcripts began.

Transcripts were randomly assigned to coders for CBCA evaluation. Each transcript was evaluated by two of the four coders. In cases of disagreement between the two coders regarding the overall credibility of the statement, the transcript was given to a third coder for credibility assessment. The final CBCA decision was based on the agreement of two of the three ratings.

Coding was done over a period of 6 months. Weekly meetings were held during which the coders' evaluations of two transcripts coded by all were jointly reviewed to ensure that the method was being properly applied, and to check on the coders' reliability over time. Any difficulties encountered with the coding were discussed.

The child's statement was initially looked at as a whole, in order to determine if it met the first three CBCA criteria (See Table 1). The statement was then evaluated on a line by line basis in order to assess presence of the remaining content criteria. The resulting data included (a) an overall score reflecting number of content criteria met, (b) a total CBCA score reflecting the overall presence and strength of the content criteria (with strength being assessed for each criterion on a 4-point rating scale: 0 = not present, 1 = questionable presence, 2 = present, and 3 = strongly present), (c) scores for each category of content criteria (reflecting both overall presence and strength of its component content criteria, and number of content criteria met within the category), and (d) scores for the presence and strength of individual content criteria. Credibility decisions were based on

the guidelines set out by Yuille (1990d), requiring the first five content criteria (in this case, excluding Criterion 4, Contextual Embedding) plus any other two criteria to be present in order to judge the statement to be valid.

It should be noted that, because of the nature of the experimental event, not all of the 19 CBCA criteria were expected to apply to the children's statements. The criteria not expected to apply were Criterion 10, Accurately Reported Details Not Understood, Criterion 17, Self-deprecation, Criterion 18, Pardoning the Perpetrator, and Criterion 19, Details Characteristic of the Offense. These criteria are linked closely with the type of offense (i.e., child sexual abuse) most commonly bringing children into contact with the criminal justice system, and are less applicable to the eyewitness situation involved in the present experimental investigation. For the purpose of the present investigation, two additional criteria were assessed (as described in section 4.6). These two criteria were included under the content category Specific Contents of the Statement. They were labelled Criterion 20, Other's Action, referring to reports of activities carried out by the other individual(s) in the event that were not specifically part of the interaction with the child, and Criterion 21, Own Action, referring to reports of the child's own activity that occurred during the target event but was not part of an interaction with the other(s). Criterion 20 and 21 are included in all CBCA analyses unless otherwise stated.

CBCA versus untrained evaluators. Twenty-five female undergraduate research assistants in the Department of Psychology, University of British Columbia, served as untrained evaluators. None of these evaluators had prior experience judging the credibility of eyewitness accounts. Further, none were familiar with the three formal types of evaluation (i.e., CBCA, amount and accuracy of detail, and Johnson/Schooler qualitative variables) employed in this study.

A subset of 30 transcripts of Grade 4 children's statements (15 randomly selected reports by children in the LE condition, and 15 randomly selected reports by children in the HC condition) were randomly ordered in five counterbalanced packages. Each package contained three LE and three HC condition transcripts. These packages were presented to untrained evaluators with a set of written instructions that included a cursory explanation of the experiment from which the transcripts were generated, and directions instructing them to try to classify each statement as credible or not credible. No information regarding the proportion of credible to noncredible transcripts was provided. Untrained evaluators were asked to indicate each of their classification decisions on a form attached to the relevant transcript. As well, they were asked to provide a confidence rating for each decision, and a written explanation of the reasons why they classified that transcript as either credible or noncredible.

The accuracy of classification decisions made by these evaluators was then compared to the accuracy of classification decisions made on the basis of CBCA for this subset of transcripts.

(2) Amount and accuracy of detail. This evaluation was conducted on the reports by Grade 4 and 2 children in the three experimental conditions using a procedure developed by Yuille, McEwan, and Kum (reported in Yuille & Cutshall, 1989). Three volunteer research assistants were trained in the application of this procedure. Training took place in three 2-hour meetings during which the application of the method was taught and several sample transcripts were coded by the group. Several 'homework' transcripts were given out at the end of each training meeting and the coders' evaluations of each were reviewed at a following meeting. Training was discontinued when it was clear that the coders were applying the method accurately and reliably.

Transcripts were randomly assigned to the three coders for amount and accuracy of detail evaluation. Each transcript was evaluated by two of the three coders. Coding was done over a period of approximately 4 months. Weekly meetings were held during which two transcripts coded by all since the past meeting were reviewed to ensure that the method was being properly applied and to check on coder reliability over time.

Coding was done by parsing each transcribed statement into a series of separate phrases (i.e., description of person, object, or action phrases). Each phrase was then analyzed to determine the number of factual details it contained. Each detail was assigned a score of 1 (e.g., "He wore a white/ cotton/ shirt" received a score of 3). In this way, scores for the total number of action details (from action phrases), person details (from descriptive phrases

relating to persons), and object details (from descriptive phrases relating to objects) were determined for each statement. Yuille and Cutshall (1989) reported that in their investigation using this method, independent analyses of the same statements by separate analyzers yielded less than 5% variance between analyzers. After determining the total number of details included in a statement, the accuracy of these details was assessed. Each detail was compared to the actual event, and on the basis of this comparison was classified as accurate, inaccurate, or unclassifiable. Unclassifiable details were those for which there was no way of assessing accuracy (e.g., self-report of the child's own internal states). The proportion of such details was very small.

(3) Johnson/Schooler qualitative variables. This evaluation procedure was applied only to the transcripts of Grade 4 children in the LE and HC conditions. The procedure involved the analysis of qualitative variables suggested by the experimental work of Johnson and colleagues (see Johnson, 1988), and Schooler et al. (1986, 1988). These variables included number of references to visual and non-visual (sound, touch, smell) sensory information, references to cognitive operations, self-references (e.g., I, me), spatial references, and verbal hedges. After developing a clear set of criteria for identifying these references, the number of references of each type were tallied for each child's statement. Three undergraduate volunteer research assistants were trained in this procedure. Training took place in three 2-hour sessions during which sample transcripts were evaluated according to the criteria.

Training was discontinued when coders were judged to be applying the procedure accurately and reliably.

Transcripts were randomly assigned to the three coders. Each transcript was evaluated by two of the three coders. Coding was done over a period of 3 months. Weekly meetings were held during which the coders' evaluations of two transcripts coded by all since the past meeting were reviewed to ensure that the method was being properly applied and to check on coder reliability over time.

Chapter 6

RESULTS

The results of this study are presented in three major sections corresponding with the three formal evaluation procedures used: (1) CBCA, including results of the comparison of classification accuracy by CBCA and untrained evaluators, (2) amount and accuracy of detail, and (3) Johnson/Schooler qualitative variables.

(1) CBCA

Interrater Reliability

Interrater reliability was assessed using Generalizability Theory (Cronbach, Gleser, Nanda, & Rajaratnam, 1972) and is reported in terms of generalizability coefficients[3]. Separate generalizability analyses were performed for the Grade 4 and the Grade 2 data. For the Grade 4 data, there were two raters(R) using 16 criteria(I) to rate the 74 transcribed statements(P). The four CBCA criteria that were judged not to have been met in any of the transcripts (i.e., Criteria 4, 10, 17, 18) were left out of these analyses. In the Grade 2 analyses, an additional criterion (i.e.

[3] Generalizability analysis, as it applies to the design of this CBCA investigation, is briefly explained by Schroeder, Schroeder, and Hare (1983) as follows:

> Since our major concern lies in reliably rank ordering individuals (the object of measurement), variance due to persons is considered universe score variance (true score variance in the classical sense), and error variance...arises from the interaction of persons with all other sources. The ratio of universe score (wanted) variance to universe score plus error (unwanted) variance reflects the generalizability coefficient (GC); it is an intraclass correlation coefficient that ranges from 0 to 1 and is interpreted in much the same way as a reliability coefficient. (p.513)

Criterion 16) was omitted from the analysis because it was not met by any of the 68 transcripts. Table 3 summarizes the generalizability coefficients, or intraclass correlation coefficients (ICCs), obtained from the raters within persons analysis. On the basis of Nunnally's (1978) guidelines for reliability, coefficients of generalizability of .70 or greater were considered acceptable in this research context. As can be seen in Table 3, the ICCs for the overall true/false judgments by raters (Grade 4 ICC = .84; Grade 2 ICC = .74) and for the 16 criteria overall (Grade 4 ICC = .69; Grade 2 ICC = .72) were sufficiently high for both the Grade 4 and Grade 2 data.

The picture is not as clear for the individual criteria. For both Grades 4 and 2, the intraclass correlations for Criteria 3, 6, 8, and 9 are clearly acceptable. The extremely low generalizability coefficients (approaching .00), as seen for Grades 4 and 2 on Criterion 1, Grade 4 Criterion 5, Grade 2 Criterion 11, and Grade 4 Criterion 20, are simply due to the extremely low between subjects variability on these criteria. In such cases, rater agreement was close to 100%. On some other variables (e.g., Grade 4 Criterion 12; Grade 2 Criteria 5, 13, 14, 15) the intraclass correlations were below acceptable levels because the judges were simply not agreeing well. It is less than desirable to have variables on which the two judges did not agree. However, closer scrutiny of the raw data reveals that such disagreements were, at least partly, accounted for by the judges' assignment of the phrase in question to different, yet related criteria. For example, disagreements on Criterion 5,

Table 3

Summary of generalizability coefficients for CBCA ratings of Grade 4 and 2 data

Variables	Grade 4 (n=74)	Grade 2 (n=68)
True/False judgment by CBCA raters[a]	.84	.74
Criteria 1	-.03*	-.01*
2	.84	.63
3	.88	.85
5	.08*	.55
6	.96	.95
7	.65	.68
8	.95	.88
9	.79	.90
11	.94	-.03*
12	.40	.64
13	.77	.54
14	.72	.45
15	.72	.52
16	.00*	--*
20	.00*	.89
21[a]	.87	.62
Total number of criteria overall[b]	.69	.72

a = Model ICC(1,4) Shrout and Fleiss, 1979
b = Design V-B Cronbach, Gleser, Nanda, and Rajaratnam, 1972
* - due to extremely low between subjects variability

Description of Interaction, for Grade 2 subjects likely reflects the fact that while some judges classified the phrase in question (e.g., "I held the screw while he took out the light bulb. He asked for it back and I gave it to him") as an interaction, others attributed the components of the phrase to Criterion 20, Own Action (e.g., I held the screw, I gave it...), and Criterion 21, Other's Action (e.g. he took out the light bulb, he asked for it back).

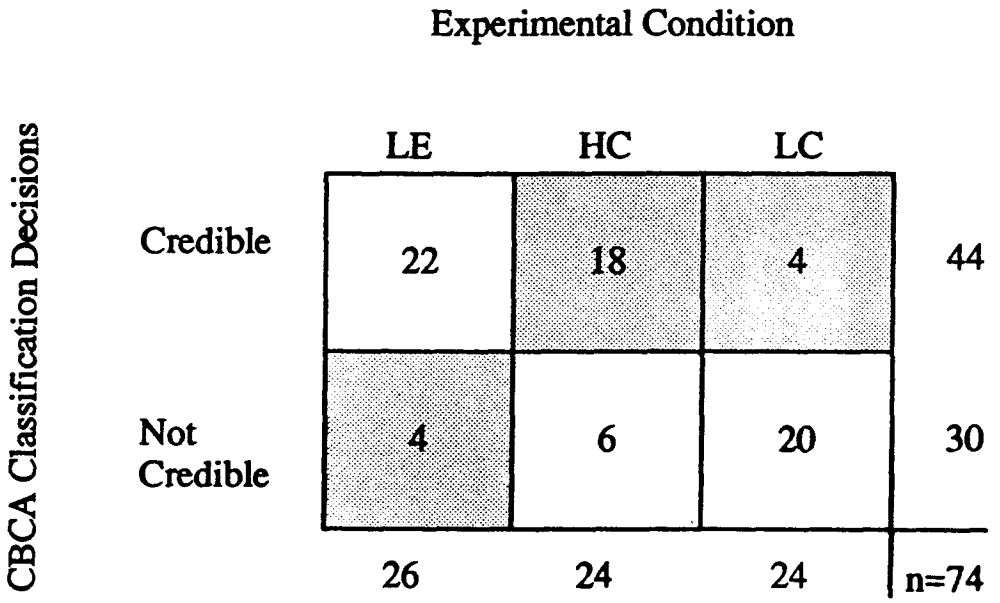
6.1 Hypothesis 1: CBCA Classification Accuracy

CBCA classifications of statements as credible or noncredible were made using Yuille's (1991d) guidelines as a formal decision rule (i.e., first five criteria plus any other two present). In this case, a transcript was considered credible if Criteria 1, 2, 3, 5, plus any other two criteria were present. Criterion 4, Contextual Embedding, was not assessed because the children were specifically asked to report only on what happened in the room while the experimenter was absent, thus eliminating any contextual embedding. Criteria 20 and 21 were not included in these analyses.

Hypothesis 1 predicted that CBCA would accurately discriminate between LE and LC transcripts across both grades. No specific predictions were made regarding CBCA classification accuracy for HC transcripts. Results are presented separately for Grade 4 and Grade 2 subjects' transcripts. Inferential comparative tests (described by Marascuilo, 1966) were used to evaluate the proportion of transcripts judged true for each experimental condition.

CBCA classification decisions for the transcripts of Grade 4 children are presented in Table 4. An analysis of variance of

Table 4. CBCA classification decisions for Grade 4 transcript



Note: Shaded blocks indicate incorrect classifications

proportions (see Marascuilo, 1966) for the Grade 4 subjects revealed that the proportion of statements judged true for each of the three experimental conditions (LE, .85; HC, .75; LC, .17) yielded a statistically significant result, $\chi^2(2,\underline{N}=74) = 47.30$, $\underline{p}<.01$. Follow-up multiple comparisons were performed (with alpha set to .05). The LC group proportion was significantly lower than the proportion for the LE and HC groups. The pair wise contrast of the LE versus HC groups was not significant. These results are summarized in Table 5.

CBCA classification decisions for the transcripts of Grade 2 children are presented in Table 6. The analysis of variance of proportions for Grade 2 subjects revealed that the proportion of statements judged true for each of the three experimental conditions (LE, .57; HC, .72; LC, .45) yielded nonsignificant results, $\chi^2(2,\underline{N}=68) = 3.72$, p = n.s. No two group proportions differed significantly from one another. See Table 7 for a summary of these results.

In summary, Hypothesis 1 was only partially supported. For Grade 4 subjects, CBCA very accurately discriminated between LE and LC transcripts. For Grade 2 subjects, CBCA did not accurately distinguish between transcripts from the two conditions.

6.2 Hypothesis 2: Number and Degree of Fulfillment of CBCA Content Criteria Across Experimental Conditions

Hypothesis 2 predicted that, across both grade levels, statements by children in the LC condition would meet the least

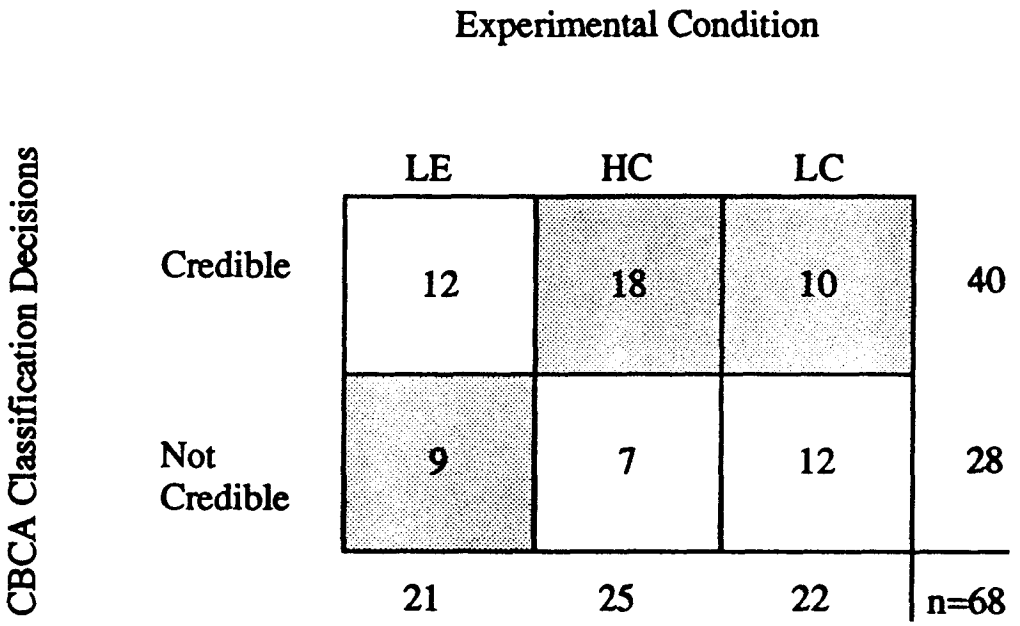
Table 5

Proportions of Grade 4 transcripts judged to be true eyewitness accounts

Transcripts	1.LE	2.HC	3.LC	Significant Mult.Compar.[a]
n	26	24	24	
Proportion judged true	.85	.75	.17	3 vs 1; 3 vs 2

[a]Overall Test: $\chi^2(2) = 47.30$. $p<.01$

Table 6. CBCA classification decisions for Grade 2 transcript



Note: Shaded blocks indicate incorrect classifications

Table 7

Proportion of Grade 2 transcripts judged to be true eyewitness accounts

Subjects	1.LE	2.HC	3.LC	Significant Mult. Compars[b]
n	21	25	22	
Proportion judged true	.57	.72	.45	none

[b]Overall Test: $\chi^2(2) = 3.72$, p = n.s.

number of CBCA criteria, statements by children in the HC condition would meet more, and statements by children in the LE condition would meet the most content criteria. A similar pattern of results was predicted with respect to the degree of fulfillment of content criteria.

The mean number of CBCA criteria met, and the mean total CBCA score (reflecting not only the number of criteria met but also the degree of fulfillment of each CBCA criteria (as rated on a 0 = not present to 3 = strongly present scale) met by Grade 4 and Grade 2 subjects in the three experimental conditions is presented in Table 8.

Two 3 (experimental condition) by 2 (grade) analyses of variance (ANOVAs) were carried out to assess differences in the number of CBCA criteria met, and the degree of fulfillment of criteria, by subjects across conditions and grades. The regression approach for decomposing sums of squares (SPSS-X Users' Guide, 1988) was used to adjust for the disproportionality in cell frequencies. Significance levels for each ANOVA were reduced on the basis of the Bonferroni inequality (i.e., alpha set to .025 [.05÷2], Howell, 1982).

The ANOVA using the number of CBCA criteria met as the dependent variable revealed a significant experimental condition by grade interaction, $\underline{F}(2,141) = 6.17$, $p<.005$ (see Figure 1). Thus, tests for the significance of the simple main effects were conducted. There was a significant simple main effect for experimental condition for Grade 4 subjects, $\underline{F}(2,136) = 25.32$,

Table 8

Mean number of CBCA criteria met and degree of fulfillment of criteria

		Number of CBCA Criteria Met		Degree of Fulfill. of CBCA Criteria	
		M	SD	M	SD
Grade 4	LE	9.54	1.49	28.54	4.43
	HC	9.35	1.68	27.79	5.28
	LC	6.46	1.68	18.90	5.15
Grade 2	LE	7.36	2.09	21.55	6.75
	HC	8.52	1.54	25.18	5.11
	LC	6.75	1.70	19.61	5.44

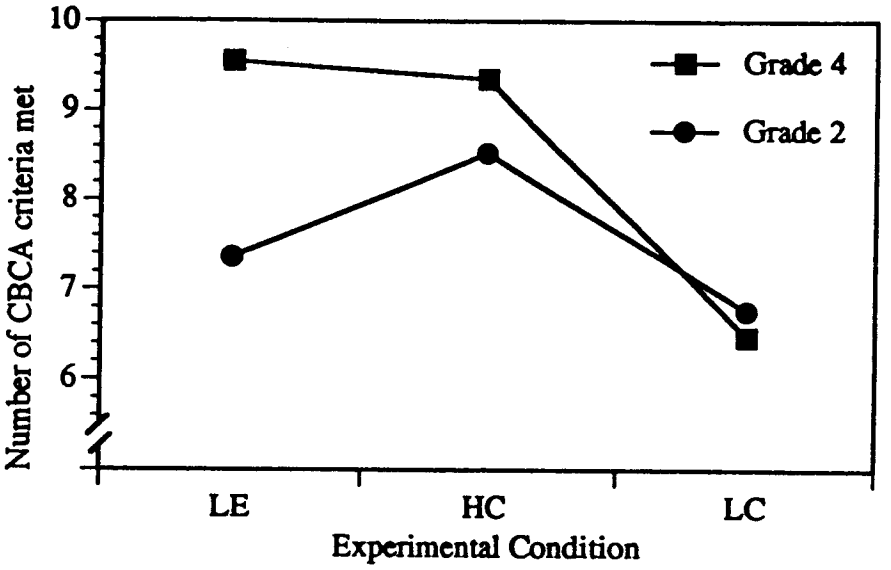


Figure 1. Number of CBCA criteria met across conditions.

$p<.001$. Similarly, there was a significant simple main effect for experimental condition for Grade 2 subjects, $F(2,136) = 6.65$, $p<.01$.

Follow-up multiple comparisons were conducted using the Tukey method adjusted for unequal n's by the Spjotvoll and Stoline procedure (1973). For Grade 4 subjects, there was a significant difference in the number of CBCA criteria met by subjects in the LE and LC conditions, with LE subjects meeting significantly more CBCA criteria than LC subjects. Similarly, there was a significant difference in the number of criteria met by subjects in the HC and LC conditions, with HC subjects meeting significantly more CBCA criteria than LC subjects. There was, however, no significant difference in the number of criteria met by subjects in the LE and HC conditions.

For Grade 2 subjects, there was a significant difference in the number of criteria met by subjects in the HC and LC conditions, with HC subjects meeting significantly more criteria than LC subjects. However, no significant differences were found in the number of criteria met by subjects in the LE and HC conditions, or more surprisingly, between the LE and LC conditions.

Next, the simple main effects for Grade were examined (see Figure 2 for illustration of Grade effects). There was a significant simple main effect for Grade for LE subjects, $F(1,136) = 19.06$, $p<.001$, with Grade 4 children meeting significantly more CBCA criteria than Grade 2 children. There were no simple main effects for Grade for HC or LC subjects.

In sum, Hypothesis 2 was only partially supported. For Grade 4 subjects, LE and HC subjects met significantly more CBCA criteria

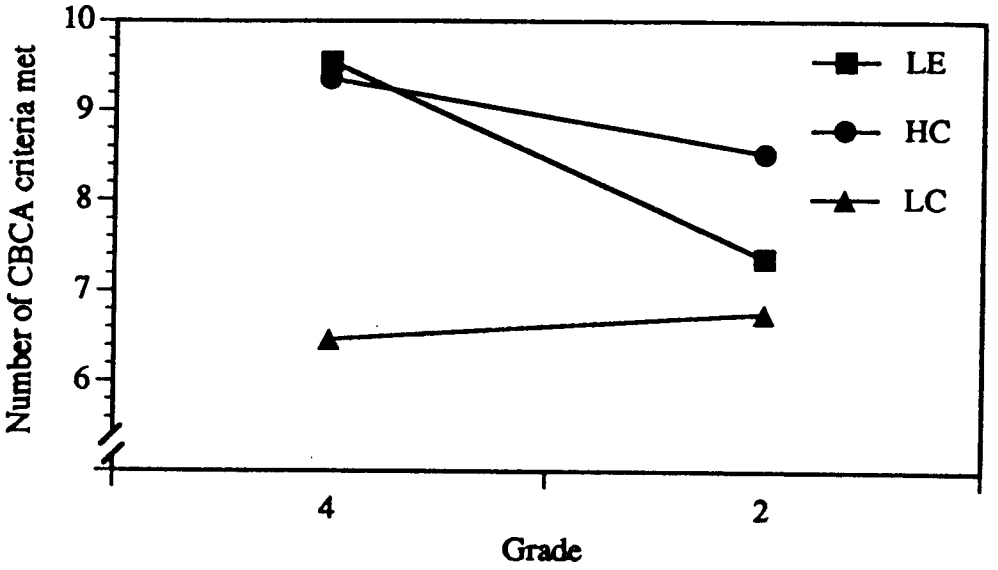


Figure 2. Number of CBCA criteria met across grades.

than LC subjects. However, LE subjects did not meet significantly more criteria than HC subjects. For Grade 2 subjects, those in the HC condition met significantly more criteria than those in the LC condition. However, the number of criteria met by LE subjects did not differ significantly from the number of criteria met by either HC or LC subjects. Grade 4 LE subjects met significantly more criteria than Grade 2 LE subjects.

The 3 (experimental condition) by 2 (grade) ANOVA using degree of fulfillment of CBCA criteria (each criterion assessed on a 0-3 scale) as the dependent variable revealed results essentially the same as above. See Appendix F for a summary of these results.

6.3 Question 3: Relative Contribution of Categories of CBCA Content Criteria to Discriminations Between Experimental Conditions

No specific predictions were made. Discriminant analyses were carried out to assess the relative contribution of categories of CBCA criteria to discriminations between the eyewitness reports by Grade 4 and Grade 2 children in the different experimental conditions.

The independent variables used in the discriminant analyses were the rationally derived categories of CBCA content criteria proposed by the developers of the CBCA (i.e., General Characteristics of the Statement--Criteria 1, 2, 3; Specific Contents of the Statement--Criteria 5, 6, 7, 8, 9; Peculiarities of Content--Criteria 11, 12, 13, 14; Motivation-Related Contents--Criteria 15, 16, 17, 18 (see Yuille, 1988). As well, Criteria 20 and 21 were included under Specific Contents of the Statement. [Of

note, analyses in which Criteria 20 and 21 were not included yielded virtually identical results]. Those criteria not met by any of the subjects were not included in the analyses (i.e., Criteria 4, 10, 17, 18, 19, plus Criterion 16 for Grade 2 subjects). Because there were different numbers of content criteria in the different categories, a mean score for each category was derived (i.e., a summed score for all criteria in the category divided by the number of criteria in the category; see Table 9 for mean scores on content categories for Grade 4 and Grade 2 subjects in the three experimental conditions).

Separate discriminant analyses were carried out for Grade 4 and Grade 2 subjects. Two sets of discriminant analyses were carried out at each grade level. In the first set, the relative contribution of the categories of content criteria to discriminations between LE, HC and LC conditions was assessed. In the second set, the aim was to achieve a clearer understanding of what was discriminating between LE and LC reports. Thus, the relative contribution of the categories of content criteria to discriminations between only the LE and LC transcripts was assessed. Any content category heavily weighted in any of these discriminant analyses was further explored by a follow-up discriminant analysis using the constitutional criteria of the category as independent variables.

The discriminant analyses carried out on the Grade 2 data yielded nonsignificant results, thus Grade 2 results are not reported. All results reported pertain to the Grade 4 data.

Table 9

Mean Content Category Scores

		Grade 4			Grade 2		
		LE	HC	LC	LE	HC	LC
General chars.	M	2.76	2.51	1.46	2.20	2.55	2.05
of statement	SD	.52	.68	.71	.87	.62	.77
Specif. conts.	M	1.87	1.97	1.51	1.49	1.90	1.44
of statement	SD	.41	.42	.56	.54	.62	.60
Peculiarities	M	1.35	1.45	.95	1.04	1.19	.90
of content	SD	.30	.35	.34	.49	.39	.41
Motivation-	M	.90	.83	.48	.19	.45	.23
related content	SD	.68	.69	.60	.37	.54	.37

1. Contribution of content categories to discriminations between the three experimental conditions (LE, HC, LC). The analysis using the four content categories as predictors of membership in the LE, HC, and LC experimental conditions yielded a highly significant discriminant function [$\chi^2(8) = 52.54$, $p<.0001$]. This discriminant function accounted for 94.66% of the discrimination. As can be seen from the coefficients presented in Table 10, this discriminant function suggests that the primary content category distinguishing between experimental conditions is General Characteristics of the Statement. Also contributing to discriminations between experimental conditions is Peculiarities of Content. Finally, Specific Contents of the Statement contributed minimally to the discrimination. Motivation-Related Contents contributed next to nothing to the discrimination. The second discriminant function was nonsignificant.

The first discriminant function correctly classified 48 of the 74 subjects (66.22%) according to group membership. More specifically, it correctly classified 16 of the 26 LE condition subjects (61.5%), 13 of the 24 HC condition subjects (54.2%), and 20 of the 24 LC condition subjects (83.3%).

Since General Characteristics of the Statement was very heavily weighted in the discriminant analysis, a follow-up discriminant analysis on the criteria comprising this category (i.e. Criteria 1, 2, 3) was carried out. This discriminant analysis yielded a highly significant discriminant function [$\chi^2(6)= 55.79$, $p<.0001$]. This function, dominated by Criterion 3, Sufficient Detail, accounted for 95.55% of the discrimination. Again, the

Table 10

Discriminant analysis of content categories for LE, HC, and LC experimental conditions

---

1st Disc. Funct: $\chi^2(8) = 52.538$, $p < .0001$, % Discrimination = 94.66

2nd Disc. Funct: $\chi^2(3) = 3.87$, $p > .1$

Standardized Normalized Discriminant Function Coefficients

	Function 1	Function 2
General Characteristics	.844	-.501
Specific Contents	.208	-.087
Peculiarities of Content	.494	.847
Motivation-related Contents	.05	-.158

---

second discriminant function did not significantly discriminate between groups (see Table 11).

This first discriminant function correctly classified 46 of the 74 (62.16%) Grade 4 children according to group membership. More specifically, it correctly classified 21 of the 26 LE condition subjects (80.8%), 5 of the 24 HC condition subjects (20.8%), and 20 of the 24 LC condition subjects (83.3%).

2. Contribution of content categories to discriminations between Grade 4 LE and LC experimental conditions. A discriminant function analysis was performed using the four content categories as predictors of membership in the LE and LC experimental conditions. There was a highly significant discriminant function [$\chi^2(4) = 41.31$, $p<.0001$]. This discriminant function suggests that the primary variable distinguishing between groups is General Characteristics of the Statement. As in the three group discriminant analysis, Peculiarities of Content contributed moderately to discriminations between conditions. The other two categories of content criteria contributed next to nothing to the discrimination (See Table 12).

This discriminant function correctly classified 44 of the total sample of 60 LE and LC transcripts (88%) according to group membership. Classification success was approximately equal for both experimental groups, with 23 of the 26 LE subjects (88.5%) and 21 of the 24 LC subjects (87.5%) being correctly classified.

A follow-up discriminant analysis was carried out using the criteria comprising General Characteristics of the Statement, which

Table 11

Discriminant analysis of General Characteristics of the Statement for LE, HC, and LC conditions

---

1st Disc. Funct: $\chi^2(6)=55.793$, $p<.0001$, % Discrimination = 95.55

2nd Disc. Funct: $\chi^2(2)= 3.525$, p = n.s., % Discrimination = 4.45

Standardized Normalized Discriminant Function Coefficients

	Function 1	Function 2
Criterion 1	-.304	.813
Criterion 2	-.068	.501
Criterion 3	.95	.297

Table 12

Discriminant analysis of content categories for LE and LC experimental conditions

Discrim. Funct: $\chi^2(4) = 41.03$, $p<.0001$

Standardized Normalized Discriminant Function Coefficients	
General Characteristics	.877
Specific Contents	.140
Peculiarities of Content	.460
Motivation-related Contents	.015

was heavily weighted in the discriminant function. Since Criterion 1, Coherence, was met by all subjects in both the LE and LC conditions (i.e., no within or between groups variance), it could not be used in the discriminant analysis. The remaining two criteria combined to define a significant discriminant function [$\chi^2(2) = 38.292$, $p<.0001$]. This function was dominated by Criterion 3, Sufficient Detail (See Table 13).

This discriminant function correctly classified 42 of the total sample of 50 LE and LC transcripts (84%) according to group membership. More specifically, 22 of the 26 (84.6%) LE transcripts, and 20 of the 24 (83.3%) LC transcripts were correctly classified.

In summary, only the discriminant analyses on Grade 4 data yielded significant results. Although the three condition discriminant analysis resulted in a significant discriminant function, its overall rate of correct classification (66.22%) was not particularly impressive. The two condition discriminant analysis, on the other hand, correctly classified 88% of the LE and LC transcripts. Both discriminant analyses carried out on the Grade 4 data found General Characteristics of the Statement to be the primary content category distinguishing among conditions. Follow-up discriminant analyses on the criteria comprising this content category revealed that Criterion 3, Sufficient Detail, contributed the most to making discriminations between conditions.

Table 13

Discriminant analysis for General Characteristics of the Statement for LE and LC conditions

Disc. Funct: $\chi^2(2) = 38.292$, $p<.0001$	
Standardized Normalized Discriminant Function Coefficients	
Criterion 2	.012
Criterion 3	.999

6.4 Question 4: Differences in Individual CBCA Content Criteria Across Experimental Conditions

No specific predictions were made. Visual inspection of the mean scores on each criterion (i.e., as rated on a 4-point scale, where 0 = not present, 3 = strongly present) for children in the three experimental conditions (See Figures 3 and 4) suggests that some criteria likely do discriminate between conditions. Analyses were carried out to investigate which criteria differed significantly across conditions. A 3 (experimental condition) by 2 (grade) multivariate analyses of variance (MANOVA) was carried out using as dependent variables the mean scores for the 10 CBCA criteria manifesting non-zero variances among subjects (i.e., Criteria 2, 3, 5, 6, 8, 12, 13, 14, 15, 21).

There was a significant overall condition by grade interaction, F (2,136) = 2.23, $p<.01$. As well, there was a significant condition effect, F (2,136) = 5.19, $p<.001$, and a significant grade effect, F (1,136) = 3.87, $p<.001$. Thus, follow-up 3 (condition) by 2 (grade) ANOVAs, adjusted for disproportionality of cell frequencies using the regression approach (SPSS-X Users' Guide, 1988), were conducted for each of the 10 CBCA criteria. The main effect for condition is reported for criteria that had a significant condition effect but not a significant condition by grade interaction (i.e., Criteria 6, 8, and 20). For each of the criteria on which there was a significant condition by grade interaction (i.e., Criteria 2, 3, 13, and 15), the univariate F-value for the interaction is reported, followed by the simple main

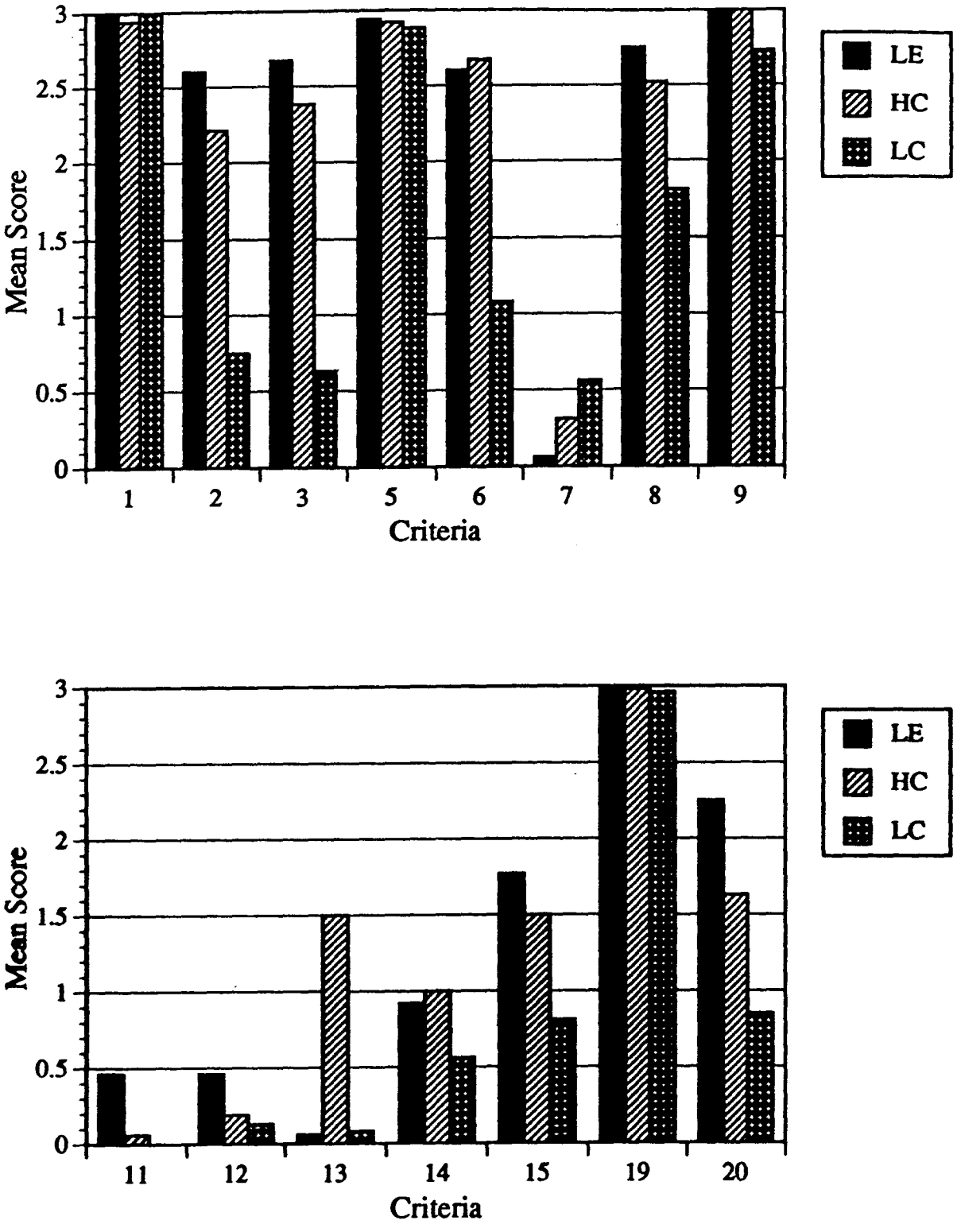


Figure 3. Mean scores of Grade 4 subjects on the 15 CBCA criteria

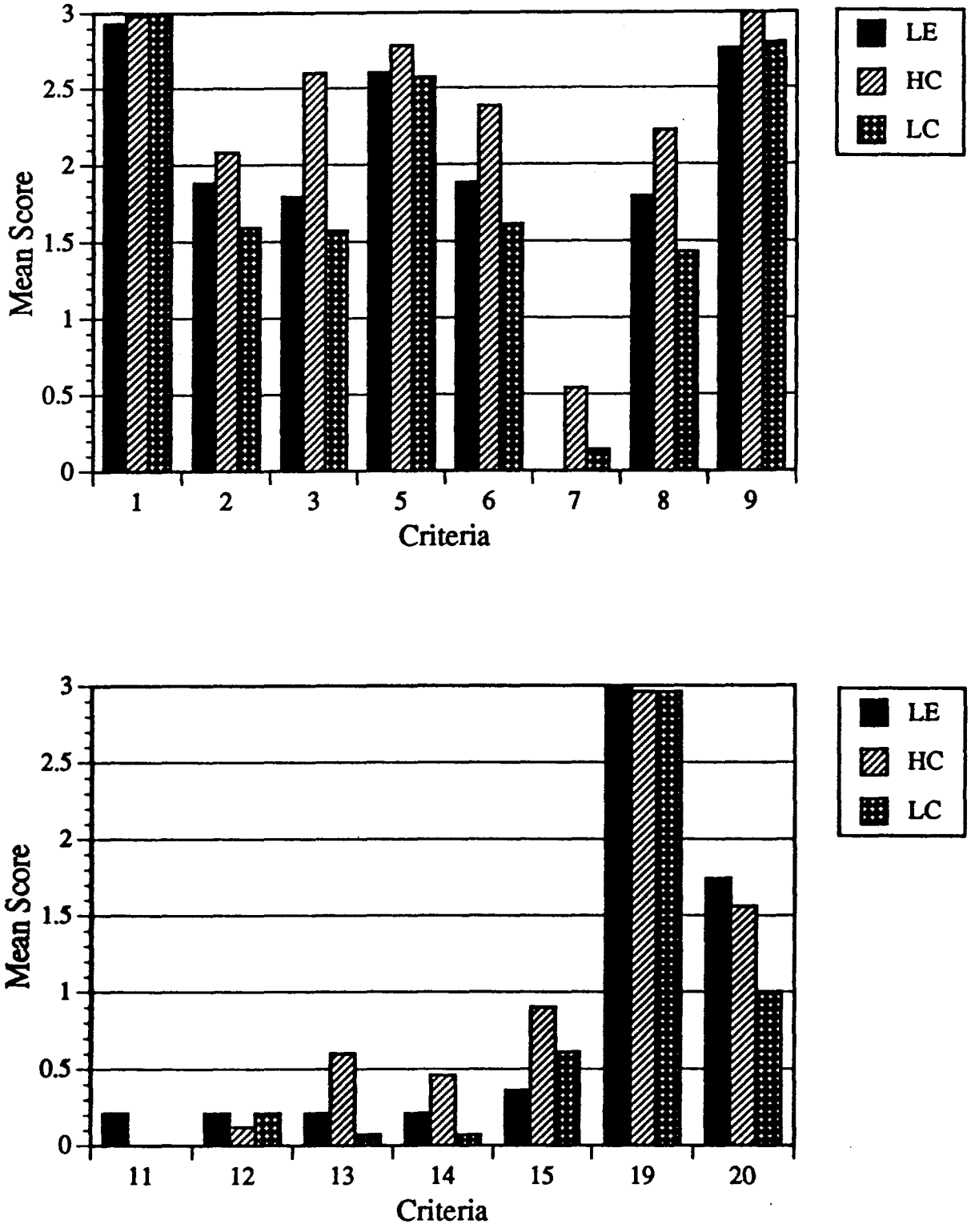


Figure 4. Mean scores of Grade 2 subjects on the 15 CBCA criteria

effect test of significance for condition. All significant main effects and simple main effects were followed up with multiple comparisons using the Tukey method, adjusted for unequal n's by the Spjotvol and Stoline procedure (1973).

Criteria with Significant Condition Effect, No Interaction

Criterion 6, Reproduction of Conversation. There was a significant main effect for condition on Criterion 6, $F(2,136) = 11.53$, $p<.001$. Follow-up multiple comparisons revealed that while subjects in the LE and HC conditions did not differ significantly from one another, subjects in both of these conditions scored significantly higher on Criterion 6 than LC subjects.

Criterion 8, Unusual Details. There was a significant main effect for condition on Criterion 8, $F(2,136) = 4.91$, $p<.01$. Follow-up multiple comparisons revealed that subjects in the LE and HC conditions did not significantly differ from each other, but scored higher on this criterion than LC subjects.

Criterion 21, Own Action. There was a significant main effect for condition on Criterion 21, $F(2,136) = 8.05$, $p<.001$. Follow-up multiple comparisons revealed no differences between the LE and HC subjects, but did reveal that both LE and HC subjects scored significantly higher than LC subjects on this criterion.

Criteria for which there was a Significant Condition by Grade Interaction

Criterion 2, Spontaneous Reproduction. There was a significant condition by grade interaction on Criterion 2, $F(2,136)$

= 5.68, $p<.005$. As well, there was a significant simple main effect for experimental condition for Grade 4 subjects, $F(2,136) = 18.70$, $p<.001$, but not for Grade 2 subjects. Follow-up multiple comparisons on the Grade 4 results revealed that subjects in the LE and HC conditions did not differ from each other on Criterion 2, but received significantly higher scores on this criterion than subjects in the LC condition.

Criterion 3, Sufficient Detail. There was a significant condition by grade interaction on Criterion 3, $F(2,136) = 8.76$, $p<.001$. The simple main effect for experimental condition was significant for both Grade 4 subjects, $F(2,136) = 26.77$, $p<.01$, and Grade 2 subjects, $F(2,136) = 6.28$, $p<.01$. Similar to the results for Criterion 2, multiple comparisons on the Grade 4 data revealed that subjects in the LE and HC conditions did not differ from each other on Criterion 3, but scored significantly higher than subjects in the LC condition. For the Grade 2 subjects, those in the HC condition received significant higher scores than LE and LC subjects. There was no significant difference between LE and LC subjects.

Criterion 13, Attribution of Other's Mental State. There was a significant condition by grade interaction on Criterion 13, $F(2,136)=6.57$, $p<.005$. The simple main effect for experimental condition was significant for Grade 4 subjects, $F(2,136) = 28.21$, $p<.001$, but not for Grade 2 subjects. Follow-up multiple comparisons on the Grade 4 data revealed that HC subjects received significantly higher scores on Criterion 13 than did subjects in the

LE or LC conditions. Subjects in the LE and LC conditions did not differ significantly on Criterion 13.

Criterion 15, Admitting Lack of Memory. There was a significant condition by grade interaction on Criterion 15, $F(2,136)=3.31$, $p<.05$. However, the simple main effect for experimental condition was not significant for either Grade 4 or Grade 2 subjects.

In sum, a significant condition effect was found for three criteria (i.e., Criteria 6, 8, and 21). On each, LE and HC subjects did not differ significantly from one another, but scored significantly higher than LC subjects. A condition by grade interaction was found for four criteria (i.e, Criteria 2, 3, 13, and 15). Simple main effects tests and follow-up multiple comparisons revealed that for Grade 4 subjects, Criteria 2, 3, and 13 differed across conditions. For all but Criterion 13, LE and HC subjects scored significantly higher than LC subjects. Interestingly, on Criterion 13, HC subjects scored significantly higher than either LE or LC subjects. For Grade 2 subjects, significant group differences were found only on Criterion 3, where HC subjects received higher scores than LE or LC subjects.

6.5 Hypothesis 5: CBCA-Trained versus Untrained Evaluators' Classification Decisions

The accuracy of credible/noncredible classification decisions made by trained CBCA evaluators (female Psychology undergraduate research assistants) and untrained evaluators (also female

Psychology undergraduate research assistants, but not trained in CBCA) were determined for each of the 30 transcripts. The CBCA score (i.e., credible/noncredible classification decision) for each transcript was the judgment made by two of the three CBCA raters for that transcript. If the original two raters randomly assigned to a transcript agreed in their classification decision, a third rater was not used. If the original two raters disagreed, the transcript was evaluated by a third rater, and the final CBCA score was that obtained by two of the three raters. The untrained evaluator score for each transcript was simply the classification decision reached by at least three of the five untrained evaluators.

It was predicted that CBCA would result in more accurate classification of LE transcripts than would untrained evaluators. No specific predictions were made regarding the classification of HC transcripts by CBCA, but untrained evaluators were expected to poorly classify HC transcripts.

CBCA Results. Fifteen of the 30 transcripts (50%) were correctly classified as credible or noncredible. Thus, overall, the CBCA evaluators' performance on this task was at chance level. However, when LE and HC transcripts were examined separately, the pattern proved more complex. Twelve of the 15 LE transcripts (80%) were correctly classified as credible, but only 3 of the 15 HC condition transcripts (20%) were correctly classified as noncredible.

Untrained Evaluator Results. Overall, 15 of the 30 transcripts (50%) were correctly classified as credible or noncredible. Thus, similar to CBCA evaluators, the untrained evaluators' performance on this task was at chance level. Unlike the CBCA results though, when LE and HC transcripts were examined separately, the picture did not change. Seven of the 15 LE transcripts (47%) were correctly classified as credible, and 8 of the 15 HC condition transcripts (53%) were correctly classified as noncredible.

Overall performance of CBCA versus untrained evaluators. The overall 'hit rate' of CBCA evaluation (i.e. 50% accuracy) relative to untrained evaluators' judgments (50% accuracy) does not require statistical analysis to demonstrate that both are performing at chance level and do not significantly differ from one another.

The performance of CBCA evaluators relative to untrained evaluators on LE and HC transcripts was assessed using chi-square analyses. The proportion of LE transcripts judged correctly did not differ between CBCA and untrained evaluators, $\chi^2(1, N=30) = 3.59$, p = n.s. Similarly, the proportion of HC transcripts judged correctly did not differ between CBCA and untrained evaluators, $\chi^2(1, N=30) = 1.89$, p = n.s.

In sum, the performance of CBCA and untrained evaluators on LE and HC transcripts combined was at chance level. Further, when performance on LE transcripts and HC transcripts were evaluated separately, there were no statistically significant differences between the rates of accurate classification by CBCA versus

untrained evaluators. Nevertheless, visual inspection of the data reveals that while untrained evaluators performed at chance level on both LE and HC transcripts, CBCA evaluators were reasonably successful at classifying LE transcripts and very unsuccessful at correctly classifying HC transcripts.

(2) Amount and Accuracy of Detail

The amount and accuracy of detail were evaluated for transcripts of Grade 4 and Grade 2 children in the three experimental conditions.

Interrater Reliability

Generalizability analyses were used to assess reliability among the three raters. Each transcript was coded by two randomly selected raters. Interrater reliability was assessed on the following variables: total number of details (i.e. person + object + action), proportion accuracy of total number of details, number of person details, proportion accuracy of person details, number of object details, proportion accuracy of object details, number of action details, proportion accuracy of action details. Generalizability coefficients, presented in Table 14, demonstrate very high intraclass correlations between raters on all variables.

6.6 Hypothesis 6: Amount and Accuracy of Detail

Means and standard deviations on all amount and accuracy of detail variables for Grade 4 and 2 subjects in the LE, HC, and LC conditions are presented in Table 15. The total amount of detail

Table 14

Generalizability coefficients for amount and accuracy of person, object and action details

	Total number of details	Accuracy of details
Total Details (Person + object + action)	.98	.88
Person	.96	.84
Object	.97	.84
Action	.97	.83

Table 15

Means and standard deviations for amount and accuracy of detail variables

Variable		Condition					
		LE		HC		LC	
		Grade 4	Grade 2	Grade 4	Grade 2	Grade 4	Grade 2
Total Number of Details	M	87.29	58.18	91.74	82.39	46.91	49.88
	SD	(26.89)	(31.79)	(26.62)	(32.15)	(19.55)	(22.87)
Proportion Accuracy of Total Detail	M	.83	.82	.80	.78	.83	.77
	SD	(.09)	(.15)	(.11)	(.10)	(.15)	(.17)
Number of Person Details	M	4.83	3.79	11.00	10.28	2.85	6.24
	SD	(2.51)	(2.76)	(3.56)	(4.28)	(3.31)	(7.09)
Proportion Accuracy of Person Detail	M	.84	.76	.79	.74	.82	.66
	SD	(.17)	(.28)	(.15)	(.19)	(.29)	(.36)
Number of Object Details	M	37.53	26.05	34.46	35.80	18.52	20.17
	SD	(17.83)	(16.40)	(11.19)	(15.56)	(8.21)	(8.61)
Proportion Accuracy of Object Details	M	.81	.85	.85	.82	.82	.80
	SD	(.09)	(.12)	(.11)	(.12)	(.16)	(.15)
Number of Action Details	M	44.93	28.35	46.28	36.31	25.54	23.48
	SD	(14.47)	(15.11)	(18.52)	(15.82)	(10.97)	(12.98)
Proportion Accuracy of Action Detail	M	.85	.85	.77	.78	.85	.85
	SD	(.11)	(.16)	(.16)	(.16)	(.12)	(.13)

(i.e., person + object + action) and proportion accuracy of this total detail across conditions and grades are presented in Figure 5.

Total amount and proportion accuracy of detail. It was predicted that the amount of detail provided by LE and HC subjects would not differ, but would be significantly higher than the amount of detail provided by LC subjects. Grade 4 subjects were expected to provide more detail than Grade 2 subjects. No differences in the accuracy of the detail provided were expected between conditions or grades.

Two 3 (experimental condition) by 2 (grade) ANOVAs, with alpha set to .025 (.05 ÷ 2) were conducted to evaluate the amount and proportion accuracy of total detail across conditions and grades. The regression approach (SPSS-X Users' Guide, 1988) was used to adjust for disproportionality in cell frequencies.

In the first ANOVA, the total amount of detail was examined. There was a significant condition by grade interaction, $F(2,134) = 4.10$, $p<.025$. Thus, tests for the significance of the simple main effects were conducted. There was a significant simple main effect for experimental condition for Grade 4 subjects, $F(2,134) = 20.05$, $p<.001$. Similarly, there was a significant simple main effect for experimental condition for Grade 2 subjects, $F(2,134) = 9.12$, $p<.01$.

Follow-up multiple comparisons were conducted using the Tukey method adjusted for unequal n's by the Spjotvoll and Stoline procedure (1973). For Grade 4 subjects, there was a significant difference in the total amount of detail given by subjects in the LE and LC conditions, with LE subjects giving more detail than LC subjects. Similarly, there was a significant difference in the

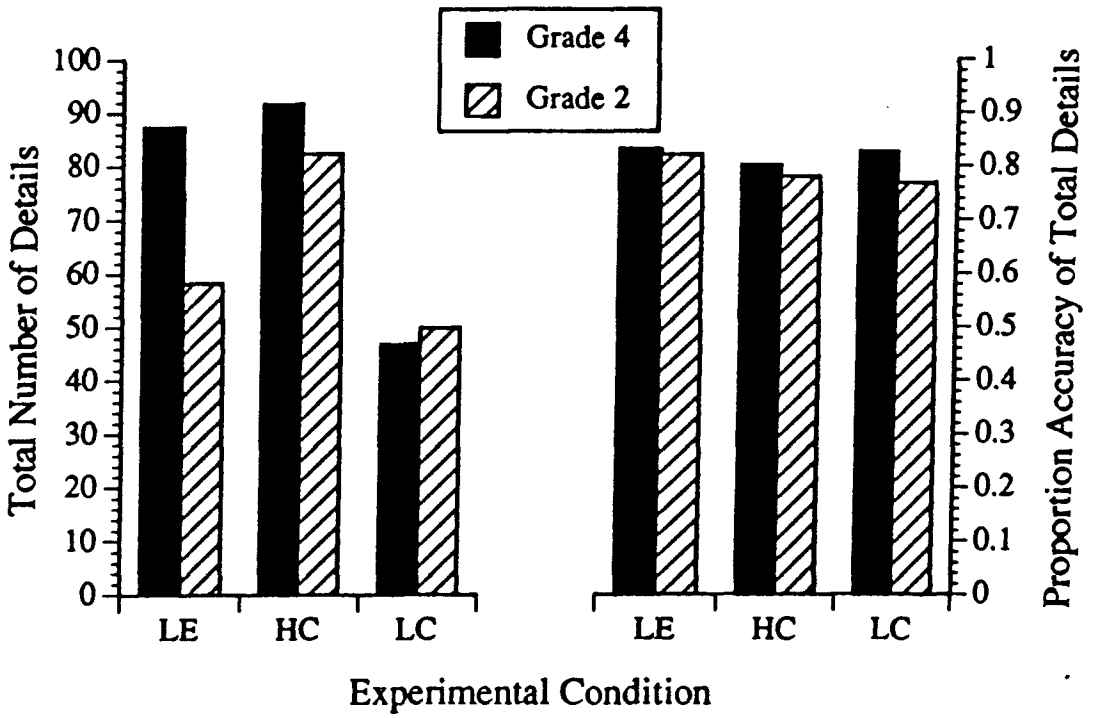


Figure 5. Total amount and proportion of accuracy of detail.

amount of detail given by subjects in the HC and LC conditions, with HC subjects reporting more detail than LC subjects. There was, however, no significant difference in the amount of detail reported by subjects in the LE and HC conditions.

For Grade 2 subjects, there was a significant difference in the amount of detail reported by subjects in the LE and HC conditions, with HC subjects giving more detail than LE subjects. Further, there was a significant difference in the amount of detail reported by HC and LC subjects, with HC subjects giving more detail than LC subjects. LE and HC subjects did not significantly differ in the amount of detail reported.

The simple main effects for grade were also examined. There was a significant simple main effect for grade for LE subjects, $F(1,134) = 13.47$, $p<.001$, with Grade 4 LE subjects giving significantly more detail than Grade 2 LE subjects. There were no simple main effects for grade for HC or LC subjects.

In the second ANOVA, the proportion accuracy of the total amount of detail was examined. There was no significant main effect for condition or grade, and no grade by condition interaction.

In sum, as predicted, Grade 4 LE and HC subjects did not differ from one another in the total amount of detail provided, but both groups gave significantly more detail than LC subjects. For Grade 2 subjects, however, this hypothesized pattern of results was not observed. Grade 2 HC subjects gave significantly more detail than either LE or LC subjects, with the latter two groups not significantly differing from one another. The hypothesized grade differences were only partially supported. Grade 4 LE subjects gave

significantly more total detail than Grade 2 LE subjects. However, there were no grade differences for HC or LC subjects. Further, as predicted, there were no significant differences in the accuracy of reported details across conditions or grades.

Amount and proportion accuracy of person, object, and action detail. In an attempt to obtain more fine-grained information about potential differences across conditions and grades in the type of details reported, a 3 (experimental condition) by 2 (grade) MANOVA was carried out with the amount and accuracy of person, object, and action details serving as dependent variables. That is, dependent variables included (a) number of person details, (b) accuracy of person details, (c) number of object details, (d) accuracy of object details, (e) number of action details, and (f) accuracy of action details. No specific predictions were made.

There was no significant multivariate condition by grade interaction, thus the multivariate main effects for condition and grade were examined. There was a significant multivariate main effect for condition, $F(2,134) = 13.44$, $p<.001$. Follow-up univariate ANOVAs on the individual dependent variables revealed a significant condition effect for total number of person details, $F(2,134) = 35.87$, $p<.001$, total number of object details, $F(2,134) = 17.75$, $p<.001$, total number of action details, $F(2,134) = 15.77$, $p<.001$, and proportion of correct action details, $F(2,134) = 3.85$, $p<.05$. Follow-up multiple comparisons were conducted using the Tukey method adjusted for unequal n's by the Spjotvoll and Stoline procedure (1973). The number of person details reported by LE and

LC subjects did not differ, but both groups gave significantly fewer person details than HC subjects. On both total number of object details and total number of action details, LE and HC subjects did not significantly differ from one another, but both gave significantly more detail than LC subjects. There were no differences between LE and LC subjects in the accuracy rate of action detail, but both of these groups gave significantly more accurate action detail than did HC subjects.

There was a significant multivariate main effect for grade, $F(1,134) = 4.07$, $p<.001$. Follow-up univariate ANOVAs on the individual dependent variables revealed only that Grade 4 subjects reported a significantly greater number of action details than did Grade 2 subjects, $F(1,134) = 14.39$, $p<.001$. There was, however, no significant difference between grades in the accuracy rate of these details. There were no other significant grade differences.

Thus, when total amount of detail and total accuracy of detail were assessed in terms of person, object, and action details, more information was obtained. HC subjects were found to provide a greater number of person details, but no greater accuracy of such details, than subjects in the other two conditions. On total number of object details, LE and HC subjects were comparable, and both groups provided more details than LC subjects. Again, there were no differences in the accuracy rate of these details across conditions. Of interest, although HC subjects provided as many action details as LE subjects, and like LE subjects, provided a significantly greater number of action details than LC subjects, the accuracy rate of action details provided by HC subjects was found to be significantly

lower than that of LE or LC subjects. Finally, Grade 4 subjects provided a greater number of action details than did Grade 2 subjects.

(3) Johnson/Schooler Variables

A number of the qualitative characteristics identified by Johnson (1988) and Schooler et al. (1986, 1988) as distinguishing real and imagined/suggested memories (i.e., number of visual and nonvisual sensory references, number of references to cognitive operations, self-references, spatial references, and verbal hedges) were investigated in the reports of Grade 4 LE and HC condition children.

Interrater Reliability

Generalizability analyses were used to assess interrater reliability among the three raters on these variables. Each transcript (of Grade 4 LE and HC subjects) was coded by two randomly selected raters. Interrater reliability was assessed on the following variables: Nonvisual sensory information, including number of sounds reported (e.g., noises heard, verbatim accounts of conversation), number of reports of touch sensations (e.g., the feel of holding screw in hand), number of smells reported, visual sensory information (including number of color, size, and shape references with regard to persons and objects), reports of cognitive operations during and following the event, number of self-references (i.e., reference to 'I', 'me', 'we'), number of spatial references (e.g., location of persons or objects in spatial context), and number of

verbal hedges (e.g., 'I think', 'it might have been ...'). Generalizability coefficients for the two raters on each of these variables are presented in Table 16. All generalizability coefficients reflect acceptable levels of rater reliability with the exception of ratings of Size - person.

6.7 Hypothesis 7: Exploratory Examination of Johnson/Schooler Qualitative Variables

Hypothesis 7 predicted that statements by children in the LE condition would contain more sensory and contextual (i.e., spatial) information, but fewer references to cognitive operations, fewer self-references, and fewer verbal hedges than statements by children in the HC condition. Means and standard deviations for the Johnson/Schooler variables are presented in Table 17.

A Hotelling's T^2 was carried out to assess differences in these qualitative characteristics between Grade 4 LE and HC subjects. There was a significant condition effect, $\underline{T}^2(1,48) = 2.83$, $\underline{p}<.05$, but follow-up univariate t-tests revealed only that subjects in the HC condition made significantly more spatial references than subjects in the LE condition, $\underline{t}(1,48) = 7.12$, $\underline{p}=<.01$. None of the other Johnson/Schooler variables were found to significantly differ across conditions. Of note, though, the difference between LE and HC subjects on the number of verbal hedges approached significance, $\underline{t}(1,48) = 3.92$, $p = .053$, with subjects in the LE condition making more hedges than those in the HC condition.

In sum, out of all the Johnson/Schooler variables investigated, only the number of spatial references differed

Table 16

Generalizability coefficients for qualitative characteristics of statements suggested by research of Johnson/Schooler

Nonvisual Sensory Information	.98
Sound	.98
Touch	.86
Smell	--*
Visual Sensory Information	.95
Color - person	.91
Color - object	.96
Size - person	.36
Size - object	.86
Shape - object	.88
Cognitive Operations	.73
Self-references	.96
Spatial references	.75
Verbal hedges	.97

* - no variance

Table 17

Means and standard deviations for Johnson/Schooler variables

	Experimental Condition			
	LE (n=26)		HC (n=24)	
	M	SD	M	SD
Nonvisual Sensory Info.	9.67	5.99	10.65	6.62
Sound	9.44	5.84	10.56	6.49
Touch	.15	.37	.08	.41
Smell	.08	.39	.00	.00
Visual Sensory Info.	13.25	7.65	14.94	4.89
Person - colour	1.44	1.20	3.83	1.40
Person - shape	.02	.10	.04	.14
Person - size	.14	.27	.54	.55
Object - colour	8.40	5.73	7.50	3.07
Object - shape	.90	1.07	.46	.66
Object - size	2.35	2.42	2.56	1.72
Cognitive References	.48	.67	.69	1.08
Self References	12.14	5.48	11.31	5.26
Spatial References	.22	.49	.72	.80
Verbal Hedges	3.15	3.20	1.69	1.77

significantly across LE and HC conditions. However, the difference was not in the direction predicted on the basis of Johnson's (cf. Johnson, 1988) earlier work.

Chapter 7

DISCUSSION

Results of the present study provide mixed support for the Undeutsch Hypothesis. In this chapter, each hypothesis and the results pertaining to it are summarized, followed by a discussion of issues related to the findings. The chapter concludes with a discussion of more general issues and directions for further research.

7.1 Hypothesis 1: CBCA Classification Accuracy

It was predicted that CBCA evaluation would accurately distinguish LE from LC transcripts across both grade levels. No particular predictions were made regarding the accuracy of discriminations between LE and HC transcripts.

Results revealed clear age-related effects. This first hypothesis was supported by the Grade 4 data. Whereas application of CBCA significantly discriminated LE from LC transcripts, it poorly discriminated LE from HC transcripts. The hypothesis was not supported by the Grade 2 data. The proportion of statements judged credible on the basis of CBCA evaluation did not significantly differ across experimental conditions.

There have been reports of the classification accuracy of CBCA from several experimental and field investigations. The reported rates of correct classification for credible and noncredible statements range from 62.3% and 77.7% (for false and true reports, respectively) in Stellar et al.'s (1988) experimental study, to

claims of 100% accuracy in Esplin et al.'s (1988) field investigation. The 85% and 83% rates of correct classifications for Grade 4 LE and LC transcripts in the present study fall in the middle of this range of results.

The variability of classification success across studies is striking. The different outcomes may reflect the fact that the type of statements subjected to CBCA differed across studies (e.g., true versus false reports of sexual abuse, reports of personally experienced versus fabricated medical experiences, reports of a witnessed staged event versus coaching). Further, outcomes may have been affected by the CBCA evaluators' degree of familiarity with the events being reported. In the present study, for instance, the CBCA evaluators were faced with many transcripts describing the same event. It is possible that over time the evaluators developed an idea of what the actual eyewitness event entailed, thus influencing their judgments regarding the credibility of later reports.

Although CBCA significantly discriminated between Grade 4 LE and LC transcripts, consideration of the classification misses reveals relatively high Type I (i.e., classification of a credible statement as noncredible) and Type II (i.e., classification of a noncredible statement as credible) error rates (i.e., 15% and 17%, respectively). In evaluating the gravity of these test misses, it is important to note that this experiment posed a particularly difficult test for CBCA, as CBCA was used in a compromised manner in a number of ways.

First, the Step-Wise Interview used to obtain the children's eyewitness reports was employed in a very restricted way. Given the

limited amount of time each child could be taken from class, the rapport building phase of the interview was very brief, and did not include elicitation of the child's reports of two irrelevant experiences. Thus, the CBCA evaluators did not have a sample of each child's reporting style (on events distinct from the experimental event of interest) by which to judge the spontaneity and sufficiency of detail in the target events description. In addition, the interview was further restricted in order to reduce the risk of children in the HC and LC conditions blatantly revealing that they had been coached. Interviewers asked for the children's eyewitness reports of the experimental events without asking any questions that would have directly addressed whether the child had personally experienced the event. In the forensic context, interviewers would almost certainly ask questions (e.g., "Did someone tell you what to say?") that would probe whether the child's report had been coached or prompted, and that would potentially lead to some children admitting that their reports were the product of an adult's coaching. Further, as described in section 5.6, the interviewer asked the children to limit their reports to the period of time beginning when the experimenter left the room and ending when she returned to the room. This instruction had the benefit of ensuring that the coached children would not begin their reports with a description of the experimenter's coaching instructions. Unfortunately, though, it also ensured that the children's reports would not place the event within the context of their day's activities, thus virtually eliminating the possibility of meeting Criterion 4, Contextual Embedding.

In addition to Criterion 4, there were a number of other CBCA criteria that were not relevant to the experimental eyewitness event and subsequent reports (e.g., Criterion 10, Accurately Reported Details Not Understood, and Criterion 17, Self-deprecation). Thus, in the present study, the statement analysis procedure was compromised because a number of criteria that could, in the forensic context, play a role in arriving at a credibility decision were simply not available to CBCA evaluators.

Further, credibility judgments made in the forensic context are based on CBCA evaluation AND information obtained through the Validity Checklist. For example, corroborative evidence (e.g., medical/physical findings, confession by the accused, etc.) would be taken into account in evaluating the child's report. The Validity Checklist was not applied (it, in fact, would have been largely inappropriate to the type of eyewitness event) in the present study, and credibility judgments were based on CBCA alone.

In sum, the compromised manner in which CBCA was used in this study likely played a role in reducing the rate of classification success for LE and LC transcripts. Nevertheless, CBCA did significantly discriminate between these transcripts. These results provide at least mildly encouraging evidence for the validity of CBCA in distinguishing between reports based on true experience and those based on 'light coaching' for children of approximately 10 years of age.

For the purpose of this investigation, LC was considered to reflect the type of coaching provided to children in the real world. There is, however, no information presently available on just what

type of coaching (e.g., amount of detail, type of detail, method of presentation, frequency of reviews) children are subjected to in the forensic context. Thus, definitive statements regarding how well LC approximated real-life coaching cannot be made. Perhaps, in the future, it will be possible to conduct research projects in which adults and children, who had coached or been coached, are interviewed regarding the nature of the coaching involved in their false statements. Until such projects have been carried out, if ever (given the improbability of finding willing subjects for such a study), the assumption that LC reflected real-life coaching will have to remain a very big, but openly acknowledged, assumption.

HC, on the other hand, was not assumed to be similar to the coaching provided to children in the forensic context. The HC script was written by individuals who were very knowledgeable about the features, originally identified by Undeutsch (1984) and, considered by CBCA to be credibility enhancing. In fact, the coaching in HC was tailored specifically to provide children with features meeting CBCA criteria. This condition was designed to permit a very challenging test of CBCA. That is, inclusion of the HC condition was expected to allow a determination of whether CBCA evaluation could be 'fooled' by procedures designed specifically to maximize the likelihood of its failure.

The results of this test indicate that by subjecting children to the CBCA-tailored coaching, and obtaining their reports immediately following the coaching, it was indeed possible to 'fool' the system. Thus, a very important boundary condition of CBCA has been demonstrated: CBCA (as used in the present study) cannot be

expected to accurately classify as noncredible testimony by children of at least 10 years of age who have been coached (i.e., deliberately trained) in features meeting CBCA criteria, and interviewed immediately after the coaching.

This finding cannot be taken as evidence that the Undeutsch Hypothesis is off the mark or that CBCA does not work in the forensic context. To reiterate, HC was designed to provide a maximally, perhaps unreasonably, difficult test for CBCA. The coaching (replete with details meeting qualitative criteria viewed as credibility-enhancing by the Undeutsch Hypothesis and CBCA) was most certainly unlike any real-life coaching. Thus, on the positive side, the misclassification of Grade 4 HC reports as credible suggests that CBCA succeeded in fulfilling its primary function, that of picking out details reflecting qualitative features hypothesized to be associated with credibility. On the negative side, though, when both live and coached accounts met the qualitative criteria postulated as *criteria of reality*, CBCA (at least as used in the present study) was not sophisticated enough to differentiate between the credible and noncredible reports.

The credibility of reports by the Grade 2 subjects was more difficult to judge with CBCA. It is not clear why the expected pattern of results was not found. The Grade 2 children provided relatively brief eyewitness accounts. The brief accounts provided by these younger children meant that the statements subjected to CBCA evaluation were shorter, therefore likely making CBCA evaluation more difficult than was the case with Grade 4 transcripts.

Similar to the Grade 4 results, the high rate of misclassification of Grade 2 HC transcripts as credible suggests that it may be possible to coach children as young as 7 to 8 years of age to present false testimony that will fool CBCA evaluation. The poor overall CBCA classification results for Grade 2 transcripts suggests that the use of CBCA with children of this age requires further evaluation. It is possible that the compromised manner in which CBCA was applied in the present study (e.g., no intra-individual standard for evaluating the sufficiency of detail, not all criteria applicable to the eyewitness event) simply made the test of CBCA impossibly difficult with the reports of these younger children. Alternatively, it may be that CBCA could have been successfully applied to statements by children of age 7 to 8 if the CBCA evaluators had a sophisticated understanding of developmental factors affecting the testimony. In the present study, CBCA evaluators were provided with a brief review of cognitive-developmental factors thought to be important to children's eyewitnessing abilities. It is possible that in spite of this review, the CBCA evaluators were simply not sufficiently attuned to developmental issues that should have been taken into account in evaluating the eyewitness recall of this younger group of children.

Finally, it may be that the system itself is not developmentally sensitive enough to permit assessments of the credibility of statements by younger children. For example, guidelines suggesting the fulfillment of seven criteria for a statement to be judged as credible (see Yuille, 1990d) may set unrealistic expectations for statements by younger children. The

finding that Grade 4 LE subjects met significantly more CBCA criteria than Grade 2 LE subjects (but there were no developmental differences in the number of criteria met for children in either of the coached conditions) supports this possibility. More research is needed to investigate the ways in which reports of real experience change as a function of age.

A number of researchers (e.g., Goodman, 1992; Gordon, Ornstein, & Schroeder, 1991; Wells & Loftus, 1991) have raised concerns about the lack of attention to developmental factors in assessing the credibility of children's eyewitness reports. Specifically, Wells and Loftus (1991) questioned "the ability of CBCA to partition individual and age-relate differences in linguistic abilities from validity-related differences" (p. 168). They suggested that the cautions by those promoting CBCA to "consider the age, experience and cognitive capacity of the witness" when applying CBCA (Raskin & Esplin, in Wells & Loftus, p. 170) are not enough to ensure that CBCA is being applied in an age-appropriate manner. Further research is needed to address the influence of specific developmental factors (e.g., age-related and individual differences in event perception, memory processes, and verbal reporting) on children's eyewitnessing abilities. The guidelines set out by the developers of CBCA for evaluating the credibility of children's statements (Yuille, 1990d) will have to be adjusted on the basis of research findings in order to make CBCA more developmentally sensitive. As well, attention must be focused on ensuring that individuals serving as CBCA evaluators are sensitive to, and able to properly assess, the role of developmental

and individual differences factors in children's eyewitness reporting. At this point, though, results of the present study suggest that extreme caution be used in applying CBCA to statements by children under 10 years of age.

7.2 Hypothesis 2: Number and Degree of Fulfillment of CBCA Content Criteria Met Across Experimental Conditions

It was predicted that there would be increments in the number of criteria met and in the degree of fulfillment of these criteria, for both grades across LE, HC, and LC conditions.

This hypothesis was only partially supported by the data. For Grade 4 subjects, LE and HC transcripts did not differ in terms of the number, or degree of fulfillment, of CBCA criteria met. However, LE and HC transcripts met significantly more criteria than LC transcripts. For Grade 2 subjects, HC transcripts met significantly more CBCA criteria than LC transcripts. No significant differences emerged in the number of criteria met by LE and HC transcripts, or by LE and LC transcripts. Of note, Grade 4 children in the LE condition met significantly more CBCA criteria than Grade 2 children in the LE condition, but there were no significant grade differences in the number of criteria met for children in either the HC or LC conditions.

These findings add to our understanding of the classification accuracy of CBCA evaluation. Yuille's suggested guidelines for categorizing a statement as likely credible (i.e., first 5 criteria plus any other 2 criteria) was used as a decision rule for classifying statements as credible/noncredible in the present study

(although Criterion 4, Contextual Embedding, did not apply in this study). This decision rule, as is the case with all suggested guidelines presently employed, is based on common sense and knowledge gained through forensic experience rather than on empirical validation. Therefore, it could be argued that the decision rule used in the present study is nonoptimal, and that CBCA may have better discriminated between conditions if a more appropriate (perhaps empirically derived) decision rule had been applied. For example, whereas the Grade 4 HC transcripts were misclassified using Yuille's guidelines as a decision rule, it may be the case that an examination of continuous scores (i.e., summed scores on all criteria) would have revealed that LE subjects met more criteria, and met them more fully, than HC subjects. The results of the comparison of continuous CBCA scores (i.e., number of criteria met and degree of fulfillment of criteria) across conditions allowed this possibility to be examined.

The Grade 4 results suggest that HC transcripts were not wrongly judged to be credible simply because they met the minimum number of criteria necessary to be classified as credible according to the decision rule used. Grade 4 HC transcripts met just as many criteria, and met them just as fully, as LE transcripts. Further, consistent with the results regarding Hypothesis 1: CBCA Classification Accuracy, this analysis based on continuous scores demonstrated clear discrimination between the LE/HC transcripts and those of children in the LC condition. It is difficult to directly compare this latter result to the findings of Esplin et al.'s (1988) field study because the two studies based CBCA total scores on a

different total number of criteria, and the degree of fulfillment of criteria was based on different rating scales. Nevertheless, it is interesting to note that the discrimination between LE/HC and LC transcripts on continuous CBCA scores in the present study is reminiscent of the clear difference in CBCA total scores for confirmed and unconfirmed cases reported by Esplin et al.

Similar to the Grade 4 findings, the comparison of continuous CBCA scores across conditions for the Grade 2 subjects demonstrated that inaccurate credibility decisions were not simply a function of inappropriate cutoffs being used in the decision-rule. Interestingly, LE subjects did not differ from subjects in either of the other two conditions in the number or degree of fulfillment of criteria. This finding, particularly the lack of differentiation between LE and LC reports, coupled with the finding that Grade 4 LE subjects met significantly more content criteria (and met them more fully) than Grade 2 LE subjects, raises the possibility that the younger children exposed to the live event had some difficulty with the basic cognitive/memory abilities required to adequately perceive, encode, retrieve, and verbally report the details of the witnessed event. This possibility is further discussed in section 7.6.

7.3 Question 3: Relative Contribution of Categories of CBCA Content Criteria to Discriminations Between Experimental Conditions

To date, there is no literature on the relative contribution of various CBCA criteria, or categories of criteria, to discriminations between live and coached/fabricated reports. Thus,

this analysis was exploratory and was initiated at a gross level, with categories of content criteria, rather than individual criteria, serving as the independent variables. No specific predictions were made.

Significant discrimination was not possible for Grade 2 subjects; no combination of content categories significantly predicted group membership. In contrast, significant results were obtained for Grade 4 subjects. In the three group discriminant analysis (i.e., relative contribution of content categories to discriminations between Grade 4 LE, HC, and LC transcripts), General Characteristics of the Statement was the primary content category distinguishing between conditions. Within this content category, Criterion 3, Sufficient Detail, played the largest role in discriminating between conditions. However, because HC transcripts were not distinguishable from LE transcripts on the basis of the CBCA criteria, these discriminant functions were not particularly successful at classifying subjects according to group membership.

A second set of discriminant analyses were carried out comparing only the Grade 4 LE and LC transcripts in order to assess whether the removal of HC transcripts would change the pattern of results. General Characteristics of the Statement (dominated by Criterion 3, Sufficient Detail) continued to play the largest role in discriminating between conditions. This time, though, the discriminant functions provided reasonably successful classification of LE and LC transcripts according to group membership.

The idiosyncratic nature of this, or any single, experimental event makes it inappropriate to make generalizations (on the basis

of these results) about which criteria or categories of criteria, are of greatest importance in discriminating between true and false/coached reports. It is possible that very different findings would be obtained in studies with different subject samples (e.g., different ages), different events, and even different application of CBCA criteria (e.g., judgments of Criterion 3 based on an intra- rather than an inter-subject standard). Zaparniuk and Yuille (1992) reported that in their pilot experimental investigation with adults, discriminant analyses revealed that Criterion 2, Spontaneity, played the largest role in discriminating between credible and noncredible reports. Clearly, generalizability of results to other experimental and field investigations is a matter for further empirical investigation.

7.4 Question 4: Differences in Individual CBCA Content Criteria Across Experimental Conditions

No specific hypotheses were made regarding which CBCA criteria would differ across conditions. In total, mean scores on 10 criteria were entered into the 3 by 2 MANOVA. Significant differences across conditions at both grade levels were found for 3 of the 10 criteria (i.e., Criterion 6, Reproduction of Conversation, Criterion 8, Unusual Details, and Criterion 21, Own Action), with LE and HC transcripts scoring significantly higher than LC transcripts in each case. For the Grade 4 subjects, three additional criteria differed significantly across conditions. On Criterion 2, Spontaneity, and Criterion 3, Sufficient Detail, LE and HC subjects again scored significantly higher than LC subjects. On Criterion

13, Other's Mental State, HC subjects scored significantly higher than LE or LC subjects. For the Grade 2 subjects, only one criterion (aside from the three for which there was a significant condition effect across both grade levels) differed significantly across conditions; HC subjects scored significantly higher than LE or LC subjects on Criterion 3, Sufficient Detail.

The results of this analysis must be viewed as a conservative test of the discriminatory power of the individual CBCA criteria. Many criteria were not included in the analysis because of an absence of variance within one or more conditions. As an example, if a criterion was met in 100% of the reports by children in one condition, and in 0% of the reports by children in another condition, this lack of within groups variance meant that this criterion could not be included in this analysis, even though it perfectly discriminated between groups. As well, the generalizability coefficients for 7 of the 10 criteria entered into this analysis indicated relatively poor agreement among raters for transcripts in at least one grade of subjects. With regard to the Grade 4 data, Criterion 5, Description of Interactions, and Criterion 12, Accounts of Subjective Mental State, two of the criteria for which no group differences emerged, had low intraclass correlation coefficients between raters. For the Grade 2 data, 6 of the 8 criteria (i.e., Criterion 2, Spontaneous Reproduction, Criterion 5, Descriptions of Interactions, Criterion 12, Accounts of Subjective Mental State, Criterion 13, Attribution of Perpetrator's Mental State, Criterion 14, Spontaneous Corrections, and Criterion 15, Admitting Lack of Memory) that did not differ across conditions

had unacceptably low intraclass correlation coefficients between raters. Since validity cannot exceed reliability, it would be unreasonable to expect these criteria to differentiate conditions. Thus, these results suggest that the validity of a number of the individual criteria were constrained by the lack of reliability among raters.

This demonstrated lack of reliability between raters on some criteria was not expected given the claims of previous investigators (e.g., Brigham & Landry, in press) that CBCA could be effectively taught to student raters in a single brief session, and the two reports of high interrater reliability on the individual criteria (Anson, 1991; Stellar et al., 1988). The results of the present study, coupled with the findings of Yuille's (in press) field study, suggest that CBCA is more difficult to apply, or to apply consistently and accurately, than the reports of these previous authors would have us believe. Nevertheless, the sufficiently high intraclass correlation between raters for the 16 CBCA criteria overall demonstrates that CBCA was applied consistently enough to look beyond reliability issues to the validity of the measure.

Before elaborating on the findings of these analyses, it is important to note that the observed group differences, or lack of differences, on individual criteria in the present study were based on eyewitness reports of an event (or coaching detailing the same event) that was highly standardized within conditions. Further, the eyewitness reports took place immediately following the witnessed event or coaching. Thus, differences in the presence or absence of specific criteria across conditions may reflect more on the

particular eyewitness event/coaching used in this study, and on the fact that long-term retention of the details was unnecessary, than on forensically relevant differences in the qualitative characteristics of credible and noncredible reports by children. Because the context of the particular event being reported is so important in determining which of the specific criteria are met (e.g., many events simply won't have an unexpected interruption or details that would necessarily be considered unusual), a thorough discussion of each criterion will not be undertaken. Instead this discussion is focused only on the criteria that are likely not specific to the idiosyncratic eyewitness event used in the present study.

With regard to General Characteristics of the Statement, Criterion 1, Coherence, was not included in the analysis because all reports were judged coherent. On Criterion 2, Spontaneity, and Criterion 3, Sufficient Detail, it appears that Grade 4 subjects in the HC condition were as able as LE subjects to report the events in a spontaneous manner and with enough detail to meet Criterion 3. The LC subjects, though, appear to have been unable to make their reports come across spontaneously, or to produce enough detail to meet Criterion 3.

For Grade 2 subjects, there were no differences across conditions on Criterion 2, Spontaneity. It is unclear if Grade 2 LC subjects were simply able to perform well and present their false testimony in a spontaneous manner, or if the expectations of CBCA evaluators were lowered for this age group of children, thus enabling even LC subjects to meet the criterion. Interestingly, the

group differences for Grade 2 subjects on Criterion 3 (i.e., LE=LC<HC), fits the emerging picture of the younger subjects in the LE condition having difficulty with one or more of the component skills required in moving from event perception to event reporting.

On these two criteria (i.e., Criterion 2, Spontaneous Reproduction, and Criterion 3, Sufficient Detail), one might expect changes in the performance of subjects in the different experimental conditions relative to each other over varying delays. For example, it may be that with a delay of two weeks instead of two minutes, LE subjects at both grade levels would be found to meet Criterion 3 to a greater extent than subjects in either of the coached conditions, as the strength of a memory is thought to be related to the extent to which the to-be-remembered event is personally significant, interesting, and directly involving (e.g., Goodman et al., 1987). Theoretically, at least, these qualities are met to a greater extent in the LE than in the HC or LC conditions.

The finding that at this brief delay, HC subjects met Criterion 3 as well as (in the case of Grade 4) or better than (in the case of Grade 2) LE subjects raises the possibility that this criterion is, in its generality, inadequately defined. This issue is further discussed in section 7.6.

It is important to note that in the forensic context, judgments of the sufficiency of details are made largely by comparing the amount of detail given regarding the event of interest with the amount of detail the child gives when talking about unrelated real-life experiences (as assessed in the early stages of the Step-Wise Interview). As previously noted, such intrasubject

comparisons could not be made in the present study. Instead, assessments of the sufficiency of details were made on the basis of the evaluators' subjective judgments of what is 'sufficient', and relative to the amount of detail provided by other children in the study. Thus, generalizability of findings regarding Criterion 3, Sufficient Detail, from this study to the forensic context is questionable.

In terms of Specific Contents of the Statement, Criterion 5, Reports of Interaction, and experimental Criterion 20, Reports of Other's Actions, were largely met in reports by children in all conditions. In contrast, there were group differences on two related criteria, Criterion 6, Reproduction of Conversation, and experimental Criterion 21, Reports of Own Action. The differences suggest that those who experienced the event gave not only general descriptions of interactions, but verbatim accounts of conversations, and tended to include descriptions of their own actions during the event. The HC subjects, who were provided with these details in the coaching, did incorporate them into their reports. However, LC subjects did not, on their own, report such details in an attempt to make their statements sound credible. Given that most eyewitness events in which a child is personally involved include these components, these results reinforce the value of Criterion 6 in helping to identify credible reports, and suggest that further consideration be given to including Criterion 21 in CBCA.

With regard to Motivation-Related Contents, there were no condition or grade effects on Criterion 14, Spontaneous Correction,

and Criterion 15, Admitting Lack of Memory. Criteria 16, 17, and 18 occurred so infrequently that they couldn't be included in the analysis. It is possible that these motivation-related criteria were not met, or did not differ across groups, because the event being reported was nontraumatic (although, theoretically, trauma should not be required for these criteria to be met). Alternatively, at least some of the criteria may have been absent because the eyewitness reports were given immediately following the event, thus because retrieval from long-term memory was not involved, making it unlikely that a child would, for example, have to admit a lack of memory or raise doubts about his/her own testimony.

Finally, proponents of CBCA have portrayed Motivation-Related Contents (such as Admitting Lack of Memory and Raising Doubts about One's Own Testimony) as reflecting honesty, and have suggested that "children who are lying will not wish to suggest doubt in their memory since the listener may then infer dishonesty" (Bekerian & Dennett, in press, p. 12). Bekerian and Dennett questioned the logic of this reasoning, and suggested instead that these motivation-related contents are likely to be rare in both truthful and fabricated accounts because children who are telling the truth will also be concerned that the listener believe their accounts and will, therefore, likely avoid making comments that they perceive as damaging to the credibility of their accounts.

Generalizability of the findings from this set of analyses to other children, events, contexts, etc. remains an empirical question. In the experimental studies conducted by Landry and

Brigham (in press) and Stellar et al. (1988), the individual CBCA criteria were found to distinguish between true and false reports to a greater extent than was found in the present study. However, their studies, too, had shortcomings, including idiosyncrasies in the type of eyewitness events reported on, very limited training of CBCA evaluators, and problematic statistical analyses. Both studies used independent t-tests to compare scores on each criteria across the two conditions, with no consideration of the likelihood of inflated experiment-wise error.

Stellar (1989) called for "more research... to overcome the present lack of knowledge about the differential validity of the various criteria of CBCA and their contribution to the overall SVA" (p. 142). Clearly, if CBCA is to become a scientifically validated instrument for credibility assessment, the differential validity of each criterion and its relative weighting in the overall decision-making process will have to be clarified. In practice, however, this state of affairs will be exceptionally difficult to achieve. The absence of clear ground-truth criteria in most forensic cases makes it very difficult to use the results of field investigations for such validation studies. Although allowing a firm knowledge of the 'truth' of the event, the results of any one experimental investigation will depend largely on the idiosyncrasies of that particular experimental event. Yet, it would be seemingly impossible to create (even to be aware of, and capable of manipulating, all the critical dimensions on which the scenarios would have to differ), and to implement (in terms of time,

personnel, and money), enough different scenarios to allow general conclusions to be drawn about the validity of individual criteria.

Practical limitations aside, serious attention should be paid to the question of whether it is worthwhile to attempt to validate the individual CBCA criteria (and to assess their contribution to the final credibility decision). Consideration of just how dependent many of the criteria are on the context and idiosyncrasies of each particular eyewitness event raises doubts as to the value of such empirical investigations. It seems likely that attempts to assess the extent to which individual criteria contribute to what is essentially a qualitative, contextually-dependent judgment are inappropriate. Future research addressing the validity of CBCA would be better directed towards evaluating the accuracy of overall CBCA judgments rather than the validity of individual criteria.

7.5 Hypothesis 5: CBCA-Trained versus Untrained Evaluators' Classification Decisions

It was predicted that CBCA evaluation would lead to a high rate of correct classifications for LE transcripts, and to a higher rate of correct classifications for these transcripts than would be achieved by untrained evaluators. No particular predictions were made regarding the classification success of CBCA on HC transcripts. Untrained evaluators were expected to poorly classify HC transcripts.

The performance of both CBCA evaluators and untrained evaluators on the LE and HC transcripts combined were at chance level (i.e., exactly 50% accuracy for both groups of evaluators).

When classification accuracy was examined separately for LE and HC transcripts, CBCA evaluators were found to be reasonably successful at classifying LE transcripts as credible (80% accuracy) and reasonably unsuccessful at classifying HC transcripts as non-credible (20% accuracy). Untrained evaluators, on the other hand, performed at approximately chance level on both LE (47% accuracy) and HC (53% accuracy) transcripts. Likely because of the small sample size (i.e., only 15 LE and 15 HC transcripts), the differences in accuracy rates for LE and HC transcripts by CBCA and untrained evaluators failed to reach statistical significance[4]. It would be worthwhile to replicate this component of the study with a larger sample of transcripts in order to determine if these group differences would hold up with a larger sample.

The results of Stellar et al.'s (1988) study hint that the observed differences may hold up with a larger sample. Stellar et al. compared credibility judgments by CBCA and untrained evaluators on 88 true and 88 false (i.e., fabricated) statements by children. Although the small number of transcripts evaluated in the present study makes comparisons between studies tenuous, it is notable that Stellar et al.'s reported rate of correct classifications by CBCA evaluators for true transcripts (77.7%) is approximately equal to the 80% rate of correct classifications obtained in the present

---

[4] Chi-square analyses were carried out in order to determine what sample size would be necessary for the difference in accuracy rates obtained for the LE and HC transcripts by CBCA and untrained evaluators to be significant. In doing these analyses, I maintained the same cell proportions and artificially increased the number of transcripts. Results indicated that for both LE and HC transcripts, the obtained cell proportions were significantly different (with alpha set at .05) when the number of transcripts was increased by 50% (i.e., n(LE) = 45 and n(HC) = 45).

study. Stellar et al. reported somewhat higher accuracy by untrained evaluators for true statements (68%) than was found in the present study (47%), but the reason for this difference is not clear. The observed rates of classification success across studies for false statements are interesting but not directly comparable. For false reports (i.e., fabrications), Stellar et al. reported 62.3% accuracy of classification using CBCA. In the present study, the false reports based on heavy coaching (i.e., with features meeting a number of CBCA qualitative criteria built into the coaching) were accurately classified only 20% of the time. It would appear that this difference across studies in CBCA classification success for false statements is due to the fact that the false reports in Stellar et al.'s study were fabrications by the children themselves therefore were unlikely to include enough features meeting criteria to be judged as credible by CBCA. In contrast, the false reports by HC subjects in the present study were based on CBCA-tailored coaching, therefore did include enough features meeting CBCA criteria to fool CBCA evaluation. When judged by untrained evaluators (i.e., evaluators not only inexperienced in making credibility judgments, but also unfamiliar with CBCA and the features postulated to reflect credibility), classification success was at approximately chance level in both studies (i.e., 47% in Stellar et al.'s study and 53% in the present study).

In the present study, the classification accuracy of both CBCA and untrained evaluators was examined at a disadvantage. As previously noted, judgments made on the basis of CBCA were handicapped by the fact that not all of the criteria were relevant

to the experimental situation, the Validity Checklist was not applied, and there was no intrasubject standard for assessing the sufficiency of detail. The untrained evaluators' assessments were handicapped by the fact that these individuals were forced to judge credibility solely on the basis of a transcript, whereas real-life judgments about the credibility of a child's report would likely rely heavily on nonverbal, or extralingual, cues (e.g., eyes, body movements, voice quality).

An attempt was made to have professionals well practiced in judging the credibility of children's reports assess the credibility of the same sample of LE and HC transcripts. This component of the study was intended to provide information regarding the classification accuracy of decisions made by these individuals relative to those made by CBCA and untrained evaluators. Unfortunately, although many packages of transcripts were distributed to judges, police officers, and social workers who had expressed interest in participating in the study, very few packages were returned, and not enough data were collected to allow the desired comparisons to be made. Comparisons of the credibility judgments made by CBCA evaluators and 'expert' credibility assessors may prove to be a fruitful direction for future research. However, more adequate tests of comparative classification accuracy might be achieved by incorporating the request for descriptions of two irrelevant episodes in the Step-Wise Interview, by videotaping the children's eyewitness reports in order to enable assessors to make full use of nonverbal/extralingual cues to deception and truthfulness, and by including more transcripts in the sample.

Finally, in evaluating the results pertaining to the classification success of CBCA (alone, and in comparison to untrained evaluators), it is important to note that the base rate of false reports in the present study was very different from the base rate of false reports in the forensic context. Whereas in the forensic context, false allegations occur with low frequency, the comparison of CBCA versus untrained evaluators in the present study was based on presentation of an equal number of LE and HC reports. Further, in the examination of the classification accuracy of CBCA, there were two false reports (a HC report and a LC report) for every one true report (the LE report). Although evaluators were instructed to judge the credibility of each transcript independently, and no information about the base rate of true and false statements was provided, the generalizability of findings to the forensic context where there is a very low base rate of false allegations is questionable. Future research could address the effects of different base rates of true and false reports on credibility decisions by manipulating the proportion of true and false reports presented to CBCA evaluators.

7.6 Hypothesis 6: Amount and Accuracy of Detail

It was predicted that children in the LE and HC conditions would provide equivalent amounts of detail overall, and that both would provide more detail than children in the LC condition. Grade 4 subjects were expected to provide more detail than Grade 2 subjects. No differences were expected in the accuracy of detail between children in the live and coached conditions, or between

Grade 4 and Grade 2 subjects. No specific predictions were made with regard to the more fine-grained analysis of person, object, and action details.

As predicted, Grade 4 LE and HC subjects gave equivalent amounts of total detail, both giving more detail than LC subjects. For Grade 2 subjects, those in the LE condition surprisingly did not differ from LC subjects in the amount of detail provided, and both gave less detail than HC subjects. The expected grade effect was present only for LE subjects, where Grade 4 subjects in the LE condition gave significantly more detail than Grade 2 LE subjects. As expected, there were no significant differences in the proportion of accurate details reported across conditions or grades.

In the Persons, Objects, Actions analysis, HC subjects gave more person details than LE and LC subjects, but there were no differences in the accuracy of reported person details across conditions. For both object and action details, LE and HC subjects gave significantly more details than LC subjects. Interestingly, although LE and HC subjects did not differ in the number of action details reported, HC subjects gave a lower proportion of <u>correct</u> action details than did either LE or LC subjects. Consistent with the grade effect for number of details reported by LE subjects, Grade 4 subjects reported a greater number of details pertaining to actions than did Grade 2 subjects. However, there were no significant differences across grades in the number or accuracy of person or object details.

Reports by other researchers (e.g., Goodman & Reed, 1986) have suggested that greater emotional involvement on the part of the

witness is associated with a narrowing of focus, with more attention paid to central aspects and less attention paid to noncentral aspects of the event. Two findings from the Persons, Objects, Actions analysis in the present study may provide support for this claim. First, the fact that HC subjects provided more details pertaining to person (e.g., hair colour/style, clothing, and accessories) than did LE subjects is consistent with an interpretation in which LE subjects are seen as being more narrowly focused on the complex series of actions (in which they were personally involved) and therefore as being less attuned to details relating to the appearance of the adult confederate. HC subjects, though, by not having the opportunity to become personally involved in the event, were likely less narrowly focused on the event per se, and were therefore able (in the interview immediately following the coaching) to present the person details provided in their earlier coaching.

Second, the higher rate of accurate action details reported by LE subjects than HC subjects may also reflect the fact that LE subjects were involved in the event and were therefore more keenly attuned to the unfolding of the event. HC subjects, lacking the opportunity for involvement, may have had difficulty tracking the complex actions involved in the event. Thus, on recall, they may have made accidental errors in their reporting of some of the actions details, and may have filled in other details to fit with their loose understanding of the course of events.

Although extremely speculative, the above interpretation of results, in terms of personally involved (i.e., LE) subjects being

more focused on action and less able than coached subjects (who were given an explicit description of the person) to elaborate on details about the person, suggests the benefit of further research focused on the evaluation of CBCA Criterion 3, Sufficient Detail. It is possible that judgments of Criterion 3 could be improved by moving from a global assessment of the 'sufficiency' of detail provided by the child witness to a more specific assessment of the sufficiency of the various types of detail (e.g., person, object, action) reported.

The remaining results discussed in this section relate to developmental considerations in children's eyewitness recall. First, the findings that (a) Grade 4 LE subjects provided more detail than Grade 2 LE subjects, and (b) Grade 4 subjects provided more action details than Grade 2 subjects, with no overall differences between grades in the proportion of accurate action details, are consistent with the larger literature documenting an age-related increase in the amount but not the accuracy of details reported (e.g, Cole & Loftus, 1987; Goodman & Helgeson, 1985; King & Yuille, 1987; Marin et al., 1979; Saywitz, 1987). These findings highlight the importance of attending to age-related differences in the amount of detail recalled when evaluating the credibility of children's reports.

Flin (1991) cautioned against drawing developmental conclusions based on a single delay interval. Both Flin (1991) and Saywitz et al. (1991) have suggested that the observed age differences in memory task performance may diminish with longer retention intervals, as the older children are likely to forget

details that initially gave them an edge over the younger children. Researchers are beginning to investigate the effects of varying retention intervals on the eyewitness recall of children (of different ages) and adults (e.g., Flin, Boon, Knox & Bull, in press; Ornstein, Gordon, & Larus, 1992; Poole & White, 1992). It would be interesting to see if the observed difference in the amount of detail recalled by Grade 4 and Grade 2 subjects in the present study would decrease with increasing retention intervals.

The finding that Grade 4 LE and HC subjects gave equivalent amounts of detail when interviewed immediately after the event or coaching was expected. It is unclear, though, whether these two groups would continue to provide equivalent amounts of detail if the retention interval between event/coaching and interview was extended. It is possible that with a delayed interview, the HC subjects' total recall would drop off to a greater extent than that of the LE subjects, because of the likely difference in strength of the original memory (Brainerd & Ornstein, 1991).

In contrast to the above result, the finding that Grade 2 LE subjects reported a smaller number of total details than Grade 2 HC subjects was <u>not</u> expected. It is, however, consistent with the findings from the CBCA analyses (i.e., the low rate of classification accuracy for all Grade 2 transcripts, the lack of difference in the number of criteria met by Grade 2 LE and LC subject). The eyewitness performance of the Grade 2 LE subjects strongly suggests that this younger group of children had a problem with some component(s) of the task demands involved in witnessing a live event and then giving an immediate eyewitness report of that

event. Unfortunately, the nature of this problem is not clear. There are a number of areas that should be considered as the possible sources of the children's apparent difficulty. The younger children may have provided little detail of the live event because of an age-related limitation in their ability to encode and then retrieve from memory the details of the event. In terms of the task of encoding, the complexity of the event may have required more attentional/processing capacity, and/or a higher level of cognitive complexity, than was available to these 7 to 8 year old children. Further, the verbal reports of these younger children may have been limited because of the children's relatively small repertoire of strategies for retrieval of information from memory. However, the above explanations do not account for the fact that Grade 2 LE subjects reported significantly less detail than children of the same age in the HC condition.

Bekerian and Dennett (in press) and Warren-Leubeker (1991) discussed the influence of factors occurring at the time of recollection on the amount of information in children's reports. They suggested that the number of details provided need not indicate what is actually in memory, rather should serve as "an index of factors affecting the report" (Bekerian & Dennett, p. 12). Warren-Leubeker (1991) elaborated, suggesting that a distinction be drawn between event memory and event reporting, because "events that children remember perfectly clearly and completely may be reported vaguely and partially, depending on the social context of the report; the child's interpretation of that context; their current

knowledge base and level of cognitive, social, event, and communicative development..." (p. 24).

In their consideration of cognitive-developmental factors related to children's eyewitnessing abilities, Turtle and Wells (1987) drew attention to the importance of functional matches between encoding and retrieval operations. They suggested that the perceptual focus of young children may restrict successful retrieval to a particular perceptual modality. Thus, if the input was perceptual, verbal retrieval cues may be ineffective. In the case of Grade 2 LE subjects, it may be imperative, not that encoding and retrieval strategies were mismatched but, that the nature of the eyewitness experience (i.e., perceptual, physical) and of the 'sharing' of that experience (i.e., verbal report) were mismatched. Younger children who perceptually/physically experienced the live event may simply have had difficulty transforming that experience (likely still very available and clear in memory) into a verbal report.

It would be very difficult, indeed, to experimentally tease out the factors critical in accounting for the Grade 2 LE subjects' performance. It is my suspicion, though, that the Grade 2 LE subjects' seemingly impoverished reports were largely due to factors relating to event reporting. Specifically, the younger children may have done a poor job of transforming their perceptual experience (perhaps still richly represented in memory) into a detailed verbal report because of their relatively unsophisticated level of verbal communication skills, and their poor judgment regarding which

details available in memory were important/relevant enough to report.

In conclusion, it is interesting to note that the comparison of total number of details across conditions was about equally successful as CBCA evaluation in distinguishing between conditions (i.e., Grade 4 LE=HC>LC; Grade 2 HC>LE=LC). Nevertheless, the assessment of amount and accuracy of detail cannot be considered as a possible method of assessing the credibility of children's reports. In the forensic context, where the events experienced by children are often very individual, it would be meaningless to compare the number of details provided by one child with reference to one event to the number of details provided by another child with reference to a different event. Further, the common lack of corroborative evidence in forensic cases makes it impossible to judge the accuracy of reported details. The findings of these analyses do, though, add to our understanding of the quantitative characteristics of children's eyewitness reports. The observed grade effects highlight the importance of further clarifying the quantitative characteristics of reports by children of different ages, and attending to and further refining our understanding of the cognitive-developmental factors accounting for these age-related differences in performance. The results of the Person, Object, Action analysis suggest the benefit of further exploration to determine which categories of detail are more or less present in true versus coached reports.

7.7 Hypothesis 7: Exploratory Examination of Johnson/Schooler Qualitative Variables

It was predicted that statements by children in the LE condition would contain more sensory and contextual information but less references to cognitive operations, less self-references, and less verbal hedges than statements by children in the HC condition.

Results did not support these hypothesized effects. The number of spatial references was the only variable that differed significantly between LE and HC reports. Although HC subjects made more spatial references than LE subjects, a result inconsistent with the direction of the difference reported by Johnson (1988), the actual means for each group on this variable were so small as to make this statistically significant difference practically meaningless.

Reasons for this failure to find the differences expected on the basis of Johnson's (1988) and Schooler et al.'s (1986, 1988) research are not clear. However, it may simply be that the qualitative features explored in this analysis were inappropriate for distinguishing between LE and HC reports. The type of qualitative features identified by Johnson, and by Schooler et al. as distinguishing between memories for perceived and imagined/suggested events are based on the premise that memories for the two types of events are formed through different processes. Memories for perceived events, being externally generated, are thought to involve more perceptual processing, therefore to include more sensory and contextual information than memories for imagined events. Memories for imagined/suggested events, being internally

generated, or formed through thought processes, are expected to include more information about the individual's cognitive and metamemory processes (Schooler et al., 1986).

According to this formulation, it is clear that children's memory for the live event in the present study can be considered to be an externally generated memory (in this case, based on perceptual experience). The origins of HC subjects' memories for the coached events are less clear. Although the children were encouraged to imagine the scenario as the coach described it, there is no way of knowing if the children actually imagined the event, or just listened to the coach with the aim of later reporting the material presented. Thus, children's memory for the HC event cannot confidently be considered to have been internally generated. It seems more likely that, although memory for the coached event was not based on perceptual experience, such a memory would be better classified as externally generated (i.e., produced and delivered by someone external to the child's own thought processes). Thus, if the event memories for both LE and HC condition subjects were in fact externally generated, it would make sense that the qualitative features proposed by Johnson, and by Schooler et al., as differentiating externally and internally generated memories failed to differentiate LE and HC reports.

This exploratory analysis represents a first attempt to investigate the relevance of Johnson, and Schooler et al.'s experimental work to the reports of children. The results do not bode well for the likelihood of these qualitative variables differing between true and 'heavily coached' eyewitness reports by

children. However, this exploratory analysis does not provide definitive evidence either for or against the potential applicability of these qualitative variables for distinguishing between credible and noncredible testimony by children. Further research would have to be conducted in order to more thoroughly assess the possibility. It is highly recommended, though, that in any such future research, attention be paid to the unique characteristics of the false 'memories' that are being compared with true memories, for descriptions of false memories are likely to differ depending on whether they were internally versus externally generated, deliberately falsified or based on inaccurate but honestly believed memorial representations, etc.

7.8 General Conclusions and Recommendations for Future Research

The present investigation experimentally tested the Undeutsch Hypothesis, which states that credible reports by children can be distinguished from noncredible reports on the basis of quantitative and qualitative features. Three general classes of quantitative and qualitative characteristics were investigated. First, the amount and accuracy of detail were compared across the three experimental conditions and two grades. In general, Grade 4 results supported the hypothesis that LE and HC subjects would provide equal amounts of total detail, both giving more detail than LC subjects. Grade 2 results were contrary to expectations, with HC subjects giving more detail than LE and LC subjects. These results highlight the importance of paying close attention to cognitive-developmental factors in assessing children's reports. Further, the findings of

the more fine-grained Person, Object, Action assessment suggest possible refinements that may be made to CBCA Criterion 3, Sufficient Detail.

Second, this study involved an exploratory examination of whether the qualitative characteristics identified by Johnson (e.g., 1988) and Schooler et al., (1986, 1988) for distinguishing between adults' memories for real versus imagined/suggested events would be of value in distinguishing between children's reports of experienced versus coached events. Results of the Johnson/Schooler analysis were disappointing. The qualitative characteristics identified by these researchers, and examined in the reports of Grade 4 LE and HC subjects, appeared to have very little discriminating power in the present context.

Finally, the primary purpose of this investigation was to provide a stringent test of the validity of CBCA for distinguishing between credible and noncredible (i.e., coached) reports by children. From this investigation, we learned that it is indeed possible to fool CBCA. When coaching provided many details meeting CBCA criteria (i.e., HC), the majority of the older group of subjects reported these details, and CBCA--by accurately identifying the presence of these details--wrongly classified the statements as credible. Although it is important to know that the system can be fooled, the forensic relevance of this finding is questionable. Unlike the artificial situation of the present experiment, most adults in the forensic context would not know how to provide CBCA-tailored coaching. Further, children would rarely be interviewed immediately after the eyewitness event/coaching. Therefore, even if

provided with features meeting CBCA criteria, it is doubtful that they would be able to report them after a lengthy delay. More important, a good forensic interviewer would ask questions likely to lead to the undoing of children's false testimony (e.g., by pursuing critical inconsistencies, or prompting an admission by the child that the report was false).

The second, and more forensically-relevant, finding of this study was that evaluation by CBCA led to a reasonably high rate of classification accuracy (84%) for LE and LC reports by the Grade 4 children. This 84% success rate is particularly impressive in light of the fact that it was achieved using CBCA in isolation (i.e., no Validity Checklist), without the benefit of all criteria potentially applying to the statement, and in spite of CBCA having been applied to transcripts of fairly restricted eyewitness interviews. One might expect an even higher success rate in more naturalistic contexts, where interviews are less restricted, all CBCA criteria are potentially applicable, and other factors (e.g., motivation to disclose, medical evidence, behavioural indicators, etc.) can be taken into account in reaching the final credibility judgment.

The Grade 2 results (i.e., that CBCA did not distinguish between the true and coached reports) are less satisfactory, and raise serious questions about the applicability of CBCA to the reports of younger children. Future research will have to focus on developmental differences in children's eyewitnessing abilities and performance, and specifically on making CBCA more developmentally sensitive.

Thus, although the results of this study add to our understanding of the usefulness of CBCA, CBCA remains a system that has not been adequately empirically validated. This unsatisfactory state of affairs leaves researchers and forensic psychologists with a philosophical/practical dilemma regarding whether or not CBCA should be used to decide upon the credibility of children's eyewitness reports in the forensic context. At one extreme, it could be argued that since credibility decisions using CBCA are fallible, the system should not be used at all. A focus on the potential misuses of CBCA (e.g., over-reliance on CBCA evaluation in the courtroom and in making decisions regarding removing a child from the home, terminating visitation rights, etc.) would seem to promote taking such a position. However, a more reasoned consideration of the evidence thus far leads to a different conclusion.

To date, CBCA is the only credibility assessment procedure that has been subjected to any empirical testing. Some studies have suggested that it is better at identifying credible reports than are lay judges (Stellar, 1989). Thus, CBCA appears to hold more promise than any available alternative. Demonstrations of limitations to its use as an objective, quantitative test of credibility, or to its application to younger children, are important, but do not necessitate discarding the system. Rather, such demonstrations suggest that the proper use of CBCA needs to be clarified.

Undeutsch, who originally described the *criteria of reality*, clearly stated that the criteria were well suited for guiding experts in carrying out subjective qualitative assessments of the

credibility of eyewitness testimony. The developers of CBCA attempted to objectify the criteria and create guidelines that would allow CBCA to be used as a quasi-objective test of credibility.

The present study was intended to test the reliability and validity of CBCA with eyewitness reports by children. Results indicated that CBCA evaluation fell short of serving as a quasi-objective test of credibility. In fact, a major strength of this experimental investigation was that the highly controlled nature of the study made it possible to identify problems with using CBCA as a psychometric instrument. On the basis of the results of this study, together with other research findings to date (see Stellar, 1989; Yuille, 1991, in press), it appears that CBCA is not suited to being used as an objective, quantitative test of credibility. Instead, consideration of research findings leads to the conclusion that CBCA would be more appropriately used as a system for organizing the facts in a case in order to assist in forming an expert opinion regarding a statement's credibility. As such, a credibility opinion would not be based on whether or not a child's report meets a standardized (rationally derived) decision rule. Rather, such an opinion would be formed by carefully evaluating the qualitative impressions arrived at through the application of CBCA, in conjunction with consideration of the circumstances of the particular case (e.g., the child's developmental level, other information obtained through the Validity Checklist). Further, it is recommended that the use of CBCA in the court system be limited to expert testimony aimed at educating the triers of fact about "the contents of statements that are consistent or

inconsistent with accounts derived from actual experience", with the "final judgment of the validity of the statement [being] reserved for the trier of fact" (Raskin & Esplin, 1991, p. 175).

Some research questions arising directly out of the present study include the following: Would CBCA classification efficiency be better if (a) the evaluations were carried out by CBCA experts? (b) the nonexpert trained CBCA evaluators were preselected on the basis of a criterion such as critical thinking ability? and (c) the contents being evaluated were less restricted (e.g., inclusion of videotaped statements, noncompromised Step-Wise Interview)? Further, would the results (e.g., classification accuracy, presence/absence of various criteria) be different with: (a) different retention intervals? (b) different base rates of true and false reports? or (c) different types of false reports (e.g., deliberate deception as in the present study versus unintentional mis-remembering brought about by errors in memory or suggestion)? Finally, it is critical to assess the accuracy of CBCA classifications relative to judgments made by professional credibility assessors unfamiliar with CBCA.

The controlled experimental research described in this thesis is a necessary component of research attempts to evaluate CBCA. However, the nontraumatic and idiosyncratic nature of the events and coaching in experimental studies such as the present one make it very difficult to generalize results to other experimental situations or to the forensic context. Yuille (in press) pointed out that "research can only address all of the [relevant] issues if a variety of research procedures are employed. Field research

should be coordinated with experimental work to provide converging operations on the issues" (p. 12). Clearly, field research has limitations as well, but "in combination with controlled research it produces a firmer foundation for application of findings" (Yuille & Wells, 1991, p. 125). The results of the present controlled experimental study will soon be related to the results of Yuille's (in progress) field research project. Through this type of labour-intensive and time-consuming coordinated research approach, we will eventually be able to refine CBCA to make it a maximally effective component of the credibility assessment process, and to define the specific circumstances under which it can be confidently and effectively applied. At this point in time, though, the assessment of CBCA is still taking place. Until further testing is completed, CBCA should be viewed as one approach to credibility assessment that has clinical support but limited empirical validation.

REFERENCES

Anson, D.A. (1991, August). Children's statements of sexual abuse: Reliability of Criteria-Based Content Analysis. Unpublished masters thesis, University of Utah, Salt Lake City.

Bala, N. (1989, February). Double victims: Child sexual abuse and the Canadian criminal justice system. Paper presented at the Canadian Bar Association - Ontario, Annual Education Institute, Criminal Law Program, Toronto.

Bekerian, D.A. & Dennett, J.L. The truth in content analysis of a child's testimony. Unpublished manuscript.

Benedek, E.P. & Schetky, D.H. (1985). Allegations of sexual abuse in child custody and visitation disputes. In D.H. Schetky & E.P. Benedek (Eds.), Emerging issues in child psychiatry and the law (pp. 145-156). New York: Bruner/ Mazel.

Bottoms, B. & Goodman, G. (1989, August). Children's testimony for a stressful event: Improving children's reports. Paper presented at the 97th annual convention of the American Psychological Association, New Orleans.

Brainerd, C. & Ornstein, P.A. (1991). Children's memory for witnessed events: The developmental backdrop. In J. Doris (Ed.), The suggestibility of children's recollections: Implications for eyewitness testimony (pp. 1-20). Washington: American Psychological Association.

Canada Evidence Act, Bill C-15. June 23, 1987.

Ceci, S.J. (1991). Some overarching issues in the children's suggestibility debate. In J. Doris (Ed.), The suggestibility of children's recollection: Implications for eyewitness testimony (pp. 1-9). Washington: American Psychological Association.

Ceci, S.J., Ross, D.F., & Toglia, M.P. (1987). Age differences in suggestibility: Narrowing the uncertainties. In S.J. Ceci, D.F. Ross, & M.P. Toglia (Eds.), Children's eyewitness memory (pp. 79-91). New York: Springer-Verlag.

Ceci, S.J., Ross, D.F., & Toglia, M.P. (Eds.). (1989). Perspectives on children's testimony. New York: Springer-Verlag.

Ceci, S. J., Toglia, M.P., & Ross, D.F. (Eds.). (1987). Children's eyewitness memory. New York: Springer-Verlag.

Clarke-Stewart, A., Thompson, W., & Lepore, S. (1989, April). Manipulating children's interpretations through interrogation. Paper presented at the meeting of the SRCD, Kansas City, MO.

Cohen, R.L. & Harnick, M.A. (1980). The susceptibility of child witnesses to suggestion. Law & Human Behavior, 4, 201-210.

Cole, C.B. & Loftus, E.F. (1987). The memory of children. In S.J. Ceci, M.P. Toglia, & D.F. Ross (Eds.). Children's eyewitness memory (pp. 178-208). New York: Springer-Verlag.

Cronbach, L.J., Gleser, G.C., Nanda, H., & Rajaratman, N. (1972). The dependability of behavioral measurements. New York: J. Wiley & Sons.

Cutshall, J. (1985). Eyewitness characteristics and memory: An in situ analysis. Unpublished master's thesis. University of British Columbia, Vancouver.

Dale, P.S., Loftus, E.F., & Rathbun, L. (1978). The influence of the form of the question on the eyewitness testimony of preschool children. Journal of Psycholinguistic Research, 7, 269-277.

Davies, G. (in press). Research on children's testimony -- implications for interviewing practice. In C. Hollin and K. Howells (Eds.), Clinical approaches to sex offenders and their victims. Chichester, England: J. Wiley & Sons.

Davies, G., Flin, R. & Baxter, J. (1986). The child witness. The Howard Journal of Criminal Justice, 25, 81-99.

Davies, G.M., Stephenson-Robb, Y., & Flin, R. (1988). Tales out of school: Children's memory for an unexpected event. In M.M. Gruneberg, P.E. Morris, and R.N. Sykes (Eds.), Practical Aspects of Memory: Vol 1. Current Research and Issues (pp. 122-127). Chichester, England: J. Wiley & Sons.

Davies, G.M., Tarrant, A., & Flin, R. (1989). Close encounters of the witness kind: Children's memory for a simulated health inspection. British Journal of Psychology, 80, 415-429.

DeAngelis, T. (1989). Controversy marks child witness meeting. The APA Monitor, 20,(9), 8-9.

de Young, M. (1986). A conceptual model for judging the truthfulness of a young child's allegation of sexual abuse. American Journal of Orthopsychiatry, 56, 550-559.

Dent, H.R. & Stephenson, G.H. (1979). An experimental study of the effectiveness of different techniques of questioning mentally handicapped child witnesses. British Journal of Clinical Psychology, 18, 41-51.

Doris, J. (Ed.). (1991). The suggestibility of children's recollection: Implications for eyewitness testimony. Washington: American Psychological Association.

Duncan, E.M., Whitney, P., & Kunen, S. (1982). Integration of visual and verbal information in children's memory. Child Development, 53, 1215-1223.

Easterbrook, J.A. (1959). The effect of emotion on cue utilization and the organization of behavior. Psychological Review, 66, 183-201.

Ekman, P. (1985). Telling Lies: Clues to deceit in the marketplace, politics, and marriage. New York: W.W. Norton and Company.

Esplin, P.W., Boychuk, T., & Raskin, D.C. (1988, June). A field validation study of criteria-based content analysis of children's statements in sexual abuse cases. Paper presented at The NATO-Advanced Study Institute on Credibility Assessment, Maratea, Italy.

Finkelhor, D. (1984). How widespread is child sexual abuse? Children Today, 3, 18-20.

Fisher, R.P., Geiselman, R.E., & Amador, M. (1989). Field test of the cognitive interview: Enhancing the recollection of actual victims and witnesses of crime. Journal of Applied Psychology, 74, 1-6.

Flavell, J.H. (1985). Cognitive Development (2nd ed.). Englewood Cliffs, N.J.: Prentice-Hall.

Flin, R. (1991). Commentary: A grand memory for forgetting. In J. Doris (Ed.) The suggestibility of children's recollection: Implications for eyewitness testimony. (pp. 21-23). Washington: American Psychological Association.

Flin, R., Boon, J., Knox, A., & Bull, R. (in press). The effect of a five month delay on children's and adults' eyewitness memory. British Journal of Psychology.

Gardner, R. (1987). The sexual abuse legitimacy questionnaire: An instrument for differentiating between bona fide and fabricated sexual abuse allegations in children. In R.A. Gardner. The parental alienation syndrome and the differentiation between fabricated and genuine child sex abuse. New Jersey: Creative Therapeutics.

Gelfand, D.M. & Raskin, D.C. (1988). Guilty, until proven innocent: The child protectors. Contemporary Psychology, 33, 28-29.

Goetze, H.J. (1980). The effect of age and method of interview on the accuracy and completeness of eyewitness accounts. Unpublished doctoral dissertation, Hofstra University, Hempstead, NY.

Goodman, G.S. (1992, May). Commentary: Present status of Statement Validity Analysis. Paper presented at the NATO Advanced Study Institute on The child witness in context: Cognitive, social and legal perspectives, Lucca, Italy.

Goodman, G.S. (1991). Commentary: On stress and accuracy in research on children's testimony. In J. Doris (Ed.), The suggestibility of children's recollection: Implications for eyewitness testimony (pp.77-82). Washington: American Psychological Association.

Goodman, G.S. (1984a). Children's testimony in historical perspective. Journal of Social Issues, 40(2), 9-32.

Goodman, G.S. (1984b). The accuracies and inaccuracies of children's eyewitness reports. In D.C. Bross (Ed.), Multidisciplinary advocacy for mistreated children, National Association of Counsel for Children.

Goodman, G., Aman, C., & Hirschman, J. (1987). Child sexual and physical abuse: Children's testimony. In S.J. Ceci, M.P. Toglia, and D.F. Ross (Eds.). Children's eyewitness memory (pp. 1-23). New York: Springer-Verlag.

Goodman, G.S. & Clarke-Stewart, A. (1991). Suggestibility in children's testimony: Implications for sexual abuse investigations. In J. Doris (Ed.). The suggestibility of children's recollections: Implications for eyewitness testimony (pp. 92-105). Washington: American Psychological Association.

Goodman, G.S. & Helgeson, V.S. (1985). Child sexual assault: Children's memory and the law. University of Miami Law Review, 40, 181-208.

Goodman, G.S. & Reed, R.S. (1986). Age differences in eyewitness testimony. Law and human behavior, 10, 317-332.

Gordon, B.N., Ornstein, P.A., & Schroeder, C.S. (1989, August). Children's testimony in sexual abuse cases: Implications of prior knowledge and interview procedures. Paper presented at the 97th annual convention of the American Psychological Association, New Orleans.

Gordon, C.L. (1985, November). False allegations of abuse in child custody disputes. The Massachusetts Family Law Journal, 54-56.

Green, A.H. (1986). True and false allegations of sexual abuse in child custody disputes. Journal of the American Academy of Child Psychiatry, 25, 449-456.

Hala, S., Chandler, M., & Fritz, A.S. (1991). Fledgling theories of mind: Deception as a marker of 3-year old's understanding of false belief. Child Development, 62, 83-97.

Howell, D.C. (1982). Statistical methods for psychology. Boston: Duxbury Press.

Hughes, M. & Grieve, R. (1980). On asking children bizarre questions. In M. Donaldson, R. Grieve, & C. Pratt (Eds.), Early childhood development and education, (pp. 104-114). Oxford: Basil Blackwell.

Johnson, M.K. (1988). Reality Monitoring: An experimental phenomenological approach. Journal of Experimental Psychology: General, 117, 390-394.

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. & Raye, C. (1981). Reality monitoring. Psychological Review, 88, 67-85.

Johnson, M.K. & Suengas, A.G. (1989). Reality monitoring judgments of other people's memories. Bulletin of the Psychonomic Society, 27, 107-110.

Jones, D.P.H., & McGraw, J.M. (1987). Reliable and fictitious accounts of sexual abuse to children. Journal of Interpersonal Violence, 2, 27-45.

Jones, D.P.H. & Seig, A. (1987). Child sexual abuse allegations in custody or visitation disputes. In B. Nicholson (Ed.) Sexual abuse allegations in child custody and visitation disputes. Washington, D.C.: American Bar Association.

King, M.A. & Yuille, J.C. (1986). The child witness. Canadian Psychological Association Highlights, 8, 25-27.

King, M.A. & Yuille, J.C. (1987). Suggestibility and the child witness. In S.J. Ceci, M.P. Toglia, & D.F. Ross (Eds.), Children's eyewitness memory (pp. 24-35). New York: Springer-Verlag.

Landry, K.L. & Brigham, J.C. (in press). The effect of training in content-based criterion analysis on the ability to detect deception in adults. Law and Human Behavior.

Lindsay, S. (1992, March). Memory source monitoring and eyewitness suggestibility. Paper presented at the American Psychology - Law Society Biennial Meeting, San Diego.

Lindsey, D.S. & Johnson, M.K. (1987). Reality monitoring and suggestibility: Children's ability to discriminate among memories from different sources. In S.J. Ceci, M.P. Toglia, & D.F. Ross (Eds.), Children's eyewitness memory (pp. 92-121). New York: Springer-Verlag.

Loftus, E.F. (1975). Leading questions and the eyewitness report. Cognitive Psychology, 7, 560-572.

Loftus, E.F. & Davies, G. (1984). Distortions in children's memory. Journal of Social Issues, 40 (2).

Loftus, E.F. & Greene, E. (1980). Warning: Even memory for faces may be contagious. Law and Human Behavior, 4, 323-334.

MacFarlane, K. & Krebs, S. (1986). Techniques for interviewing and evidence gathering. In K. MacFarlane & J. Waterman (Eds.), Sexual Abuse of Young Children: Evaluation and treatment (pp. 67-100). New York: Guilford Press.

Marascuilo, L.A. (1966). Large-sample multiple comparisons. Psychological Bulletin, 65, 280-290.

Marin, B.V., Holmes, D.L., Guth, M., & Kovac, P. (1979). The potential of children as eyewitnesses. Law and Human Behavior, 4, 295-306.

Munsterberg, H. (1908). On the witness stand. New York: Doubleday.

Nunnally, J.C. (1978). Psychometric Theory (2nd ed.). New York: McGraw Hill.

Ornstein, P.A., Gordon, B.N., & Larus, D.M. (1992). Children's memory for a personally experienced event: Implications for testimony. Applied Cognitive Psychology, 6, 49-60.

Parker, J.F., Haverfield, F., & Baker-Thomas, S. (1986). Eyewitness testimony of children. Journal of Applied Social Psychology, 16, 287-302.

Pear, T.H. & Wyatt, S. (1914). The testimony of normal and mentally defective children. British Journal of Psychology, 3, 388-419.

Peters, D.P. (1991a). The influence of stress and arousal on the child witness. In J. Doris (Ed.), The suggestibility of children's recollections: Implications for eyewitness testimony (pp.60-76). Washington, D.C.: American Psychological Association.

Peters, D.P. (1991b). Commentary: Response to Goodman. In J. Doris (Ed.), The suggestibility of children's recollections: Implications for eyewitness testimony (pp. 86-91). Washington, D.C.: American Psychological Association.

Peters, D.P. (1987). The impact of naturally occurring stress on children's memory. In S.J. Ceci, M.P. Toglia, & D.F. Ross (Eds.), Children's eyewitness memory (pp.122-141). New York: Springer-Verlag.

Piaget, J. & Inhelder, B. (1974). The child's construction of quantities: Conservation and atomism. New York: Basic Books.

Poole, D.A. & White, L.T. (1992, May). Two years later: Effects of question repetition and retention interval on the eyewitness testimony of children and adults. Paper presented at the NATO Advanced Study Institute on The child witness in context: Cognitive, social, and legal perspectives, Lucca, Italy.

Quinn, K.M. (1988). Children and deception. In R. Rogers (Ed.), Clinical assessment of malingering and deception (pp. 104-119). New York: Guilford Press.

Raskin, D.C. & Esplin, P.W. (1991). Commentary: Response to Wells, Loftus, & McGough. In J. Doris (Ed.), The suggestibility of children's recollection: Implications for eyewitness testimony (pp. 172-176). Washington: American Psychological Association.

Raskin, D.C. & Yuille, J.C. (1989). Problems in evaluating interviews of children in sexual abuse cases. In S.J. Ceci, D.F. Ross, & M.P. Toglia (Eds.), Perspectives on child testimony (pp. 184-207). New York: Springer-Verlag.

Rogers, M.L. (1990). Coping with alleged false sexual molestation: Examination and statement analysis procedures. Issues in Child Abuse Accusations, 2, 57-68.

Rose, S.A. & Blank, M. (1969). The potency of context in children's cognition: An illustration through conversation. Child Development, 40, 383-406.

Rudy,L. & Goodman, G.S. (1991). Effects of participation on children's reports: Implications for children's testimony. Developmental Psychology, 27, 527-538.

Saywitz, K.J. (1987). Children's testimony: Age-related patterns of memory errors. In S.J. Ceci, M.P. Toglia, & D.F. Ross (Eds.). Children's eyewitness memory (pp.36-52). New York: Springer-Verlag.

Saywitz, K.J., Goodman, G., Nichols, E. & Moan, G. (1991). Children's memories of a physical examination involving genital touch: Implications for reports of child sexual abuse. Journal of Consulting and Clinical Psychology, 59, 682-691.

Schooler, J.W., Clark, C.A., & Loftus, E.F. (1988). Knowing when memory is real. In M.M. Gruneberg, P.E. Morris, & R.N. Sykes (Eds.). Practical aspects of memory: Current research and issues - Memory in everyday life (pp. 83-88). New York: J. Wiley and Sons.

Schooler, J.W., Gerhard, D., & Loftus, E.F. (1986). Qualities of the unreal. Journal of Experimental Psychology: Learning, memory, & cognition, 12, 171-181.

Schroeder, M.C., Schroeder, K.G. & Hare, R.D. (1983). Generalizability of a checklist for assessment of psychopathy. Journal of Consulting and Clinical Psychology, 51, 511-516.

Sheehy, N.P. & Chapman, A.J. (1982). Eliciting children's and adults' accounts of road accidents. Current Psychological Reviews, 2, 341-348.

Shrout, P.E. & Fleiss, J.L. (1979). Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin, 86, 420-428.

Sink, F. (1987). Studies of true and false allegations: A critical review. Paper presented at the Third National Family Violence Research Conference, Durham, NH.

Spjotvoll, E. & Stoline, M.R. (1973). An extension of the T-method of multiple comparisons to include the cases with unequal sample sizes. Journal of the American Statistical Association, 68, 975-978.

Sporer, S.L. (1982). A brief history of the psychology of testimony. Current Psychological Reviews, 2, 323-349.

SPSS-X Users' Guide (3rd ed.). Chicago: SPSS Inc.

Stellar, M. (1989). Recent developments in statement analysis. In J.C. Yuille (Ed.), Credibility assessment (pp.135-154). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Stellar, M., Wellershaus, P., & Wolf, T. (1988, June). Empirical validation of criteria-based content analysis. Paper presented at the NATO-Advanced Study Institute on Credibility Assessment, Maratea, Italy.

Stellar, M. & Koehnken, G. (1990). Statement Analysis: Credibility assessment of children's testimonies in sexual abuse cases. In D.C. Raskin (Ed.), Psychological methods for criminal investigations and evidence. New York: Springer-Verlag.

Stern, L.W. (1910). Abstracts of lectures in the psychology of testimony and on the study of individuality. American Journal of Psychology, 21, 270-282.

Suengas, A.G. & Johnson, M.K. (1988). Qualitative effects of rehearsal on memories for perceived and imagined complex events. Journal of Experimental Psychology: General, 117, 377-389.

Suski, L.B. (1986). Child sex abuse--An increasingly important part of child protective service practice. Protecting Children, 3, 3-7.

Turtle, J.W., & Wells, G.L. (1987). Setting the stage for psychological research on the child eyewitness. In S.J. Ceci, M.P. Toglia, & D.F. Ross (Eds.). Children's eyewitness memory (pp.230-248). New York: Springer-Verlag.

Undeutsch, U. (1989). The development of statement reality analysis. In J.C. Yuille (Ed.), Credibility assessment (pp. 101-119). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Undeutsch, U. (1984). Courtroom evaluation of eyewitness testimony. International Review of Applied Psychology, 33, 51-67.

Undeutsch, U. (1982). Statement reality analysis. In A. Trankell (Ed.), Reconstructing the past: The role of psychologists in criminal trials. (pp.27-56). Stockholm: P.A. Norsted & Sons.

Varendonck, J. (1911). Les temiognages d'enfants dans un proces retentissant. Archives de Psychologie, 11, 129-171.

Warren-Leubecker, A. (1991). Commentary: Development of event memories or event reports? In J. Doris (Ed.), The suggestibility of children's recollection: Implications for eyewitness testimony. (pp.24-26). Washington: American Psychological Association.

Wells, G.L. & Loftus, E.F. (1991). Commentary: Is this child fabricating? Reactions to a new assessment technique. In J. Doris (Ed.), The suggestibility of children's recollection: Implications for eyewitness testimony. (pp. 168-171). Washington: American Psychological Association.

Wigmore, J.H. (1909). The psychology of testimony. Illinois Law Review, 3, 399-434.

Yuille, J.C. (in press). Combining field and experimental research in the study of children's eyewitness memory. In M.P. Toglia, D. Peters, & S. Ceci (Eds.), The child witness: International research and legal perspectives.

Yuille, J.C. (1991). A pilot experimental investigation of the validity of CBCA. Unpublished manuscript.

Yuille, J.C. (1990a). The Step-Wise Interview : A protocol for interviewing children. Unpublished document.

Yuille, J.C. (1990b). Statement Validity Analysis: Content criteria for statement analysis. Unpublished document.

Yuille, J.C. (1990c). Statement Validity Analysis: Validity checklist. Unpublished document.

Yuille, J.C. (1990d). Use of the Criteria-Based Content Analysis. Unpublished document.

Yuille, J.C. (1988a). The systematic assessment of children's testimony. Canadian Psychology, 29, 247-262.

Yuille, J.C. (1988b, June). A simulation study of criterion-based content analysis. Paper presented at the NATO Advanced Study Institute on Credibility Assessment, Maratea, Italy.

Yuille, J.C. & Cutshall, J. (1989). Analysis of the statements of victims, witnesses and suspects. In J.C. Yuille (Ed.), Credibility assessment (pp. 175-191). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Yuille, J.C., Cutshall, J.L., & King, M.A. (1988). Age related changes in eyewitness accounts and photo-identification. Unpublished manuscript, University of British Columbia, Vancouver.

Yuille, J.C. & Cutshall, J.L. (1986). A case study of eyewitness memory of a crime. Journal of Applied Psychology, 71, 291-301.

Yuille, J.C. & Farr, V.L. (1987). Statement validity analysis: A systematic approach to the assessment of children's allegations of sexual abuse. The British Columbian Psychologist, Fall, 19-27.

Yuille, J.C., Hunter, R., & Harvey, W. (1990). A coordinated approach to interviewing in child sexual abuse investigations. Canada's Mental Health, 38, 14-18.

Yuille, J.C., Hunter, R., Joffe, R., & Zaparniuk, J. (in press). Interviewing children in sexual abuse cases. In G.S. Goodman & B. Bottoms (Eds.), Understanding and improving children's testimony: Clinical, developmental and legal implications. Guilford Press.

Yuille, J.C. & Kim, C.K. (1987). A field study of the forensic use of hypnosis. Canadian Journal of Behavioural Sciences, 19, 418-429.

Yuille, J.C & Wells, G.L. (1991). Concerns about the application of research findings: The issue of ecological validity. In J.L. Doris (Ed.), The Suggestibility of Children's Memory. Washington: American Psychological Association.

Zaparniuk, J. & Yuille, J.C. (1992, March). Assessing the credibility of true and false statements. Paper presented at the Biennial Meeting of the American Psychology - Law Society, San Diego, CA.

Zarazoga, M.S. (1987). Memory, suggestibility, and eyewitness testimony in children and adults. In S.J. Ceci, M.P. Toglia, & D.F. Ross (Eds.), Children's eyewitness memory (pp.53-78). New York: Springer-Verlag.

Appendix A

Parental Consent Form

THE UNIVERSITY OF BRITISH COLUMBIA

Department of Psychology
2136 West Mall

Vancouver, B.C. Canada V6T 1Y7
Telephone (604) 228-2755

Dear Parents or Guardian:

A research project is being conducted in your child's school to gain knowledge about the differences in children's memories for events in which they are participants and their memories for events which they are told about. This research is being organized and overseen by me, Dr. John C. Yuille, of the University of British Columbia.

The children who participate will be seen individually for a maximum of 30 minutes at a time deemed convenient by your child's teacher. During this time, they will either participate in, or be told about, a simple event. They will then be interviewed and asked to recall all they can about the event. Interviews will be audio taped for later assessment.

Please note that I am not interested in the performance of individual children, but rather, in the overall performance of children at varying ages. This project is in no way connected to any evaluation of your child; participants will be assigned identification numbers so that your child's name will not be used in subsequent information analysis.

You and your child's cooperation would be greatly appreciated. If you agree to allow your child to participate, please sign the attached form and have him/her return it to the school as soon as possible. Please also note that if your child does not wish to participate in this project, or at any time chooses to withdraw, he/she may do so without jeopardy to their class standing. Only those children who return signed consent forms, and are themselves willing to participate, will be interviewed.

If you have any questions or concerns about the project please do not hesitate to contact me at 228-6130 or my graduate student, Risha Joffe, at 228-5581.

Thank you for your time and considerations.

Sincerely,

John C. Yuille, Ph.D.
Professor

THE UNIVERSITY OF BRITISH COLUMBIA

Department of Psychology
2136 West Mall

Vancouver, B.C. Canada V6T 1Y7
Telephone (604) 228-2755

Child's Name: (please print) _____

Child's Grade: _____ Birthdate: _____

_____ Yes, I agree to allow my child to participate in this study.

_____ No, I do not wish my child to participate in this study.

Parent or Guardian's signature: _____

Appendix B

Development of Final Methodology

Development of Final Methodology

The Pilot Study

A pilot study was carried out on a group of 17 Grade Two students from a local elementary school. The children were divided into three experimental conditions. Children in the live condition were taken to the Memory Room and were witness to, and participants in, the live event. Children in the coached conditions were not exposed to the Memory Room or the live event. Instead, they were told (in varying degrees of detail, depending on experimental condition) about the room and the events that they were to pretend had happened in the room. All children then met with an interviewer and presented their recollection of what happened in the Memory Room.

Several findings from the pilot study were of importance in the development of the final methodology of this study. First, the pilot study was successful in that most of the children recalled a fair amount of the information presented via the live event or coaching. There were variations in the details recalled across children. However, there were no readily apparent differences in the quantity or type of detail recalled across groups. Of great importance, it did not appear that the event posed so heavy a memory load as to render the children unable to give any recollection of it. Further, there were no indications that the children were in any way traumatized by the demands of the experiment. Thus, changes to the event itself following the pilot study were minimal.

The pilot study allowed the identification of several problems requiring changes to the experimental procedures. In the pilot study, coaching was carried out by one of two female research assistants. The instructions presented to the coached children were as follows:

> We're interested in finding out how well you can act, - like, how well you can pretend that something happened to you. I'm going to tell you a story. Pretend that you're in the story and that it really happened. Later, another lady is going to ask you what you remember about the story. She knows that some kids had the story happen to them in real life, and some kids are pretending that it happened to them. But, she doesn't know which kids are telling it for real and which kids are pretending. She doesn't want to know till later; it's like a guessing game for her. When the lady asks you questions about what happened, we want you to try and tell her the story as if it happened to you in real life. So, you'll want to do your very best acting job to try and make her believe it really happened to you. Any questions? O.K. Listen carefully and try to remember as much as you can about the story. It might help to close your eyes and imagine that it's really happening to you.

The coach then read the script describing the to-be-remembered event. The event was presented in a storytelling manner with the action presented in the present tense (e.g.,. "You get introduced to a lady called _____ who is already in the room. You sit down at the desk.. ."). No attempts were made to involve the children in the script by requesting their input into the imaginary story.

When later interviewed, seven out of the 11 children in the coached conditions indicated that they had been coached by making statements such as "She told me a story about ..." or "then she said that the repairman ...", even by talking about the repairman

and child involved in the event in the third person (e.g.,. "Then they started playing..."). One child, who in fact avoided making these errors, lost track of her own tale and shifted from referring to the female experimenter as "the lady" to "mommy" by the latter part of her interview. Another child, later identified by the teacher as an English as a second language (ESL) student easily "spooked" by new activities, simply froze and failed to respond to any of the interviewer's queries. Only two children in the coached conditions succeeded in consistently presenting their recall for the event in the first person and with no blatant indications that they had been coached.

Final Methodology

Research findings by psychologists who have extensively researched children's capacity for using deceptive strategies (see Hala, Chandler, & Fritz, 1991) support the view that by Grade Two, children are capable of verbally misleading others into believing something which they themselves know to be false. Thus, given the very small proportion of children who were able to purposely mislead the interviewer in the pilot study, it appeared that the coaching was in some way inadequate for tapping children's ability to mislead others.

There are a number of plausible reasons why this may have been the case. Perhaps the coaching instructions did not make the task at hand clear to the children. It may be that the instructions did not provide an incentive great enough to motivate the children to put their best effort into the task. Further, the manner of

presentation of the event may not have been adequate for inviting the children's involvement in the story. Finally, the complicated experimental procedure (i.e.,. meeting with a female coach in one room, being asked by her to imagine meeting and interacting with another woman in another room, then being asked by the interviewer to report on what happened in the other room with the other woman) may simply have overwhelmed the children with demands that they could not keep straight.

Efforts were made to modify the coaching instructions and event presentation in order to deal with these problems. First, changes were made to the instructions to children in the coached conditions. In the pilot study, the child's cooperation was solicited by calling into question his/her talent as an actor. Based on discussion with others who have faced the problem of engaging children's cooperation in similar tasks (Hala, personal communication, 1990), the modified instructions were instead geared toward soliciting co-operation by peaking the child's delight in conspiring with the coach to play a harmless trick on the interviewer. The to-be-fooled interviewer was presented as a friend of the coach who enjoys a good practical joke and who would look favourably upon this prank.

The interviewer's debriefing of the child was modified to deal with the 'practical joke' aspect of the coached experimental conditions. Upon completion of the Step-Wise Interview, the child was told that his/her participation in the experiment was over. The interviewer explained that she was aware that some children were asked to report on true events whereas others were asked play a

trick on her by reporting on 'pretend' things. The interviewer then hazarded a guess as to the child's experimental condition (in reality, 'guessing' that the event truly happened in all but the most obvious of coached cases). Once this guessing game was completed, the interviewer explained the study to the child.

Changes were also made to the coaching script itself in an attempt to make it more personally involving for the child, in hopes of further reducing the problem of children telling the tale in the third person form. Thus, in the revised version, the to-be-imagined event was presented to the child in the past tense. The coach requested a small bit of imaginary input from the child early on in the story (i.e.,. a statement of which colour of LEGO the child added to the LEGO house), and at various points asked for the child's agreement (e.g.,. when showing a picture of the repairman, coach prompted "He looks friendly, doesn't he?"). Throughout the script, the telling of the event was peppered with phrases such as "Now pretend that...", "make believe that ...", etc.

In addition, in order to eliminate the problem of children reporting what went on in the coaching session when asked to give their recall of everything that happened in the Memory Room (e.g.,. "I went in and she told me a story"), the instructions to the child regarding what was to be recalled were made very specific, delimiting the information being sought to the specific time period beginning when the female experimenter left the room and ending when she returned to the room. For children in both the live and coached conditions, these instructions were intended to focus their recall on the period of time during which the repairman was in the room and

not on the events preceding or following the repairman's visit (the contents of which varied across conditions).

Finally, in an effort to make the event more like a real-life situation calling for eyewitness testimony, the main study differed from the pilot study in that the live event and coaching took place in the same room. The key props (e.g., lamp, LEGO, tape recorder), present in the room during the live event, were present when the coaching took place as well. The female experimenter in the live experimental condition doubled as the coach in the coached conditions. These changes were made in an effort to ensure that any emergent differences in the reports of live event versus coached condition children were due to factors related to having truly experienced or not experienced the event, and are not simply due to the child's differential familiarity with the individuals involved in the alleged event or the physical setting in which the event occurred.

Appendix C

Script (LE Condition)

Script (LE Condition)

- lead child into room and have him/her sit at desk with LEGO on it; you sit at desk with tape recorder.

"We're going to do some stuff that will help me learn more about how children remember things. It's really important that you pay close attention to everything that happens in this room. Later, you'll be asked some questions about what happened in here. To start with, we're going to build a house out of this LEGO"

- encourage child to put a piece onto existing base; you follow child's move by putting a small blue piece of LEGO onto the one (s)he put onto the base. Then, after blue piece is in place, have child put a couple more pieces down.
- then, suddenly look at watch. Say:
"Oh-oh! I have to make a phonecall. Please stay here and I'll be back as soon as I can".
- get up and walk to door, just before leaving say:
"Oh, a repairman might come in to fix this lamp. If he comes while I'm gone, you can help him out. OK? - then leave the room

- come back shortly after repairman leaves room. Say:
"Sorry I took so long". (notice tape recorder is gone)
"Hey, where's the tape recorder?" (child responds)
"Oh. I was afraid of that. He must've thought it was the school's tape recorder. It's really mine and we need it for this memory experiment. I'd better let him know that it's mine so we can get it back. I need it badly."
"I don't know which repairman was here, so I'm going to need your help. We'll need to know everything that happened while I was out of the room. How about coming and talking with my friend _____. She'll ask you to tell her about what happened in here while I was out of the room. Try really hard to tell her <u>everything</u> you remember, even the stuff you might think isn't important. OK? Thanks a lot. I <u>really</u> appreciate your help."

- lead child to interviewer and introduce them. Say to interviewer "While <u>child's name</u> was in my room, I had to go out to make a phonecall. Some stuff happened in the room while I was gone and <u>child's name</u> is the only one that can tell us about what happened. It would really help if you could talk with <u>child's name</u> and ask him/her to tell you everything that (s)he remembers about what happened in that other room. Thanks a lot.

Appendix D

Coaching Instructions and Scripts

COACHING INSTRUCTIONS

We're going to do some stuff that will help me learn more about how children remember things. Pretty soon you're going to have a chance to talk with one of my friends. Her name is _____.

Now, _____ and I love to play tricks on each other. I have a <u>really</u> good trick to play on her, but I'm going to need your help to do it. Want to hear my idea?

_____ is going to be talking with lots of kids today. Most of the kids are going to tell her about some things that really happened to them while they were in this room. The stuff that happened in this room while the other kids were in here isn't going to happen while you're in here. It'll be different with you. You get to do something that I think is more fun.

This is what you'll do. I'm going to tell you about the stuff that happened to the other kids, and you and I will do our very best to pretend that it <u>really</u> happened to you too.

It's very important that you pay close attention to everything I tell you. The trick will be that when you go talk with _____ and she asks you to tell her what happened, you'll pretend that the things that we just talked about really did happen while you were in here. So far so good? You'll try to make _____ believe that the stuff we talked about really happened. Won't it be fun if we can fool her and she thinks it really happened?!

I bet you're a good actor. To fool her, you'll want to try really hard to make her believe that the pretend stuff really, truly happened while you were in this room. Then, at the very end, after

_____ finishes talking with you, we'll find out if she thought the make believe stuff really happened. We'll tell her at the end that you were fooling her. _____ will be happy that she got to have a trick played on her.

Why don't we give it a try? I think this will be fun. (I can hardly wait). I want you to listen very carefully and imagine that what I tell you about really happened while you were in here. OK? When we finish, I'll take you to talk with _____. Listen carefully and try to remember as much as you can.

HEAVY COACH SCRIPT

Let's pretend that you and I came into this room and sat down at the desks. You sat at the desk with the LEGO on it. I sat at the desk with the tape recorder. Pretend that once we were sitting down, I said "We're going to build a house out of LEGO". Then we started putting together the pieces of LEGO to build a house. Imagine us doing that--make believe you reached over and got a piece. What colour? (let child give a colour). Ya, a ____ piece and you stuck it onto the bottom yellow piece of LEGO. Let's say I got a small blue piece and put it on top of your (colour) piece. Then you put on a couple more pieces. Can you picture us doing that?

O.K., now pretend that then I suddenly looked at my watch. I said "Oh-oh! I have to make a phonecall. Please stay here and I'll be back as soon as I can". Then I got up and walked to the door. Just before I walked out of the room, I said "Oh, a repairman might come in to fix this lamp. If he comes while I'm gone, you can help him out". Then I left the room.

Now pretend that soon after I left, there was a knock at the door. You looked to see who was there and there was a man standing at the door. Here's a picture of the man. He looks friendly, doesn't he? He's got brown eyes and dark brown hair that's really short. You can't tell from the picture, but he's quite tall. Pretend that when you saw him at the door, he was wearing black jeans, a white t-shirt, and blue runners. Over his jeans, he was wearing a big belt. The belt had pockets and places to hang tools

from. There was a hammer, and a screwdriver with a green handle, hanging from the belt. The man had a knapsack over one shoulder,. The knapsack was brown/orange, and it had black straps. Can you imagine him dressed like that? Good.

Keep pretending ... Pretend that the man standing at the door looked at you and said "Hello, I'm a repairman. I'm here to fix the lamp". Then he walked up to the desk where you were sitting and put his backpack down on the floor. He asked "What's that you're building?". You told him. He said "I love construction, but I'd better do my work".

Now, make believe that then he walked up to the big yellow lamp (point) and tried to turn it on. But, the light didn't go on. So, he reached to the top of the lamp and unscrewed a little gold cap. He turned to you and asked "Will you hold this for me, please?". Pretend that you got out of your desk and walked up to him. He passed you the little gold cap and said "thank you". Then he took the lamp shade off and put it down on the floor. He started to unscrewed the light bulb. But, all of a sudden an alarm started ringing. He seemed really surprised. He stopped unscrewing the light bulb and pulled a black stopwatch out of his pocket. He turned off the alarm. Then he went back to fixing the lamp. He took the light bulb out, and stuck it into the top of his t-shirt (gesture). Then he went back to his backpack and put the light bulb into it. He took a new light bulb out of the backpack and screwed it into the lamp. This time the light went on. He seemed happy. He said "Great!" and then he turned the light off.

Pretend that then the repairman went back to his knapsack and took something out. It was a picture. He looked at it and seemed to get kind of sad. Then he showed you the picture and said "This is my cat". Imagine that you looked at the picture and saw a cute little white kitten with blue eyes. The kitten was hanging over a pole/scratching post. The man looked at you and smiled a bit. Then he put the picture back into his knapsack.

Now pretend that then he walked over to the desk with the tape recorder on it (point) and turned the tape recorder on. The song "Puff the Magic Dragon" started playing. You both listened for a little while. Then the repairman said "Hmmm, it sounds a little fuzzy. Might need a cleaning. Let's see". He turned off the music and took the tape out. Then he got a different tape out of his knapsack and put it into the tape recorder. He pressed the record button and said to you "Copy me". Then he made sounds, like "lalalalala" and "budum-budum-budum". Each time he made a sound, he got you to make the same sound after him. Then, he said "Good, now we'll see if it needs fixing". He rewound the tape and you both listened to what you had just recorded. Then he turned it off and said "It's not working very well, but I think I can repair it".

Pretend that the repairman unplugged the tape recorder. Then he got a towel out of his knapsack. The towel was striped--red, pink, and white. It was quite torn up. He wrapped the tape recorder in the towel. Then he asked you "Will you help me fit this into my knapsack, please". So you helped him put the tape recorder into the backpack. When you finished he said "Thank you".

Then he said "Well, that's about it for now. Have a good day". He picked up his knapsack from the floor, waved to you, and walked out of the room.

Now, pretend that pretty soon after he left, I came back. I said "Sorry I took so long". Then I noticed that the tape recorder was gone. I asked "Hey, where's the tape recorder?". Imagine that you told me where it went. Then I said "Oh, I was afraid of that. He must've thought it was the school's tape recorder. It's really mine and we need it for this memory experiment. I'd better let him know that it's mine so we can get it back". Then I told you that I needed your help to find the repairman.

(end of script)

"OK, that's the end of what we're pretending happened. The fun part is about to start. Remember, _____ doesn't know that the stuff we just talked about didn't really happen. That's good, because we want to fool her into thinking it did happen in real life. When you get into _____'s room, she's going to ask you to tell her everything you remember about what happened in this room--from the time that I left to make my phonecall until the time I came back into the room.

Then, "Think you're ready to go talk with _____ now? Let's see if we can fool her into thinking that the pretend stuff really happened while you were in this room. This will be fun. You ready?"

- take child to interviewer. Introduce them.
- say to interviewer:

> While child's name was in my room, I had to go out to make a phonecall. Some stuff happened in

the room while I was gone and child's name is the only one that can tell us about what happened. It would really help if you could talk with child's name and ask him/her to tell you everything that (s)he remembers about what happened in that other room. Thanks a lot.

LIGHT COACH SCRIPT

Let's pretend that you and I came into this room and sat down at the desks. Pretend that once we were sitting down, I said that we would do things to help me learn about how children remember things. Then I told you that we'd start by building something out of LEGO. So, we put some pieces of LEGO together. Can you picture us doing that?

Pretend that then I suddenly looked at my watch. I said that I had to make a phonecall. I asked you to stay in the room. But, before I left, I told you that a repairman might come in. I said that you could help him. Then I left.

Now pretend that after I left, there was a knock at the door and a man came in. He told you that he's a repairman. He said that he was there to fix the lamp. He came up to where you were and put his backpack down. He asked you what you were doing. You told him. He said that he likes doing that too, but that he'd better do his work.

Now, make believe that then he went up to the lamp and tested it. It didn't work. So, he unscrewed something from the lamp and got you to hold it. Then he took off the lamp shade and put it down. He started to take out the light bulb, but then his alarm started ringing. So, he stopped fixing the lamp to turn off his alarm. Then he took the light bulb out and put it into his backpack. He got a new light bulb and screwed it into the lamp. This time the lamp worked. He turned the light off.

Pretend that then the repairman got a picture out of his knapsack and looked at it. He showed it to you and said it was a picture of his cat. Then he put the picture away.

Now pretend that then he went over to the tape recorder and turned it on. You listened for a while to the music that started playing. Then he said that the tape recorder probably needed a cleaning. He put in a different tape and started recording. He made sounds into the mike and got you to copy him. Then you both listened to what you just recorded. He said that he could fix the tape recorder to make it work better. So, he unplugged the tape recorder. Then he wrapped it in a towel and got you to help him put it into his knapsack. Then he thanked you and got ready to leave. He said goodbye, waved at you, and left.

Now, pretend that pretty soon after he left, I came back. I said that I was sorry for taking so long. Then I noticed that the tape recorder was gone. I asked where it went and you told me. Then I said that I was afraid that he might take it. I explained that he thought it was the school's tape recorder, but really it was mine. And I told you that I had to get it back from him. Then I told you that I needed your help to find the repairman.

(end of script)

"OK, that's the end of what we're pretending happened. The fun part is about to start. Remember, _____ doesn't know that the stuff we just talked about didn't really happen. That's good, because we want to fool her into thinking it did happen in real life. When you

get into _____'s room, she's going to ask you to tell her everything you remember about what happened in this room--from the time that I left to make my phonecall until the time I came back into the room.

Then, "Think you're ready to go talk with _____ now? Let's see if we can fool her into thinking that the pretend stuff really happened while you were in this room. This will be fun. You ready?"

- take child to interviewer & introduce them

- say to interviewer:

> While child's name was in my room, I had to go out to make a phonecall. Some stuff happened in the room while I was gone and child's name is the only one that can tell us about what happened. It would really help if you could talk with child's name and ask him/her to tell you everything that (s)he remembers about what happened in that other room. Thanks a lot.

Appendix E

Interview Script

INTERVIEW SCRIPT

"I'm going to ask you to tell me about what happened in the other room starting from when Robin left the room to make a phonecall up till the time that she came back from her call. I have a tape recorder here to tape what we say. That's so I can listen to it later instead of having to write everything down now. I have a pen and paper here too, because I don't want to interrupt you. So, if I think of questions to ask you, I'll write them down for later. O.K.?"

"Let's try out the tape recorder to make sure it works". (TURN TAPE RECORDER ON). "This is your name talking with child's first name and last initial at _____School on date, 1990 at time. Let's make sure both our voices come out clear. My birthday is on _____, when's your birthday? (child responds). "How about telling me about something you did in teacher's class today" (tie this up quickly). Rewind tape some. Listen, and if child's voice isn't clear, encourage speaking loudly and make adjustments to volume. (Don't erase the identifying information).

"Now, I want you to tell me about what happened while you were in the other room--from the time Robin left to make a phonecall until the time she came back. It's OK if you can't remember everything. But, it's really important that you tell me EVERYTHING that you CAN remember, even if you think that some of it isn't very important. Lets start with when Robin said she had to leave the room. Ready? Tell me everything you remember about what happened"

(if you have a reticent child, you can say something like "Remember, Robin left the room to go do something. What happened then?")

(FREE RECALL)

- be tolerant of pauses, silences

- if child stalls, OK to say "then what happened", or " you said (last thing child mentioned), what happened next?"

When free recall ends

- "Can you remember anything else about what happened?"

Follow up on specifics

- e.g,. " you said _____, can you tell me more about that?" or " you talked about a _____, can you tell me anything else about it/that/him/etc?"

Ask for repetition

"You've told me about what happened, but my memory's not so hot. It would be a real help to me if you would tell me about what happened again"

-OR-

"You've told me about what happened. I think I've got it all straight, but there are a few things I'm not sure of. If you wouldn't mind, I'd like you to tell me what happened one more time"

(CHILD REPEATS TESTIMONY)

- Follow up on specifics if something new comes up.

- "Thank you very much. You've been a real help".

DEBRIEFING

"Thanks, we're all finished now. You've done really well. Robin told me that most of the kids I'd talk with today would be telling me about something that happened while they were in the other room. But, Robin said that I might get to talk to some kids who were helping to play a trick on me. For a joke, these kids would be trying to make me believe that some pretend things happened, when they didn't really happen in real life. Now that we're all done, I'd like to guess--don't tell me yet--whether you were telling me about things that really happened, or whether you were kidding me. OK?"

"Let's see (ham this part up), I think you were (can vacillate back and forth a bit) telling me about something that really happened. Am I right?" (if wrong, show that you enjoyed being fooled - e.g., "Ah, you did it, you really pretended well. I thought it really happened. That was fun....").

"I'd like to tell you a little bit more about the memory experiment we're doing here today. We came to your school today because we are really interested in learning more about how children remember things."

- if child was in LE condition, say:

"You know the man that came into the other room while you were in there? He was really an actor putting on a play for you. So, he pretended to be a repairman, and he pretended that he had to take away a tape recorder and fix it. Really he brought it right back to

Robin. We had him put on that play for you so that we could look to see how kids remember things that happened to them in real life".

- if child was in HC or LC condition, say

" You were told a story about a repairman. Then when I asked you questions about what happened in the other room, you did your best acting job to pretend that make believe things really happened while you were in the other room. We were looking to see how kids remember make believe things".

- then to all:

"Some kids in your class haven't done our experiment yet. So, please don't tell them anything that happened. If they ask, just tell them to wait and you'll talk about it when everyone's had a chance to do it. That would make it more fair for everybody--the other kids in your class and us.

Your teacher will let the class know when everyone's finished being in the experiment. Then it will be O.K. to talk about it with your friends. In the mean time, of course you can talk about it with your parents or your teacher. Any questions? " (answer questions).

Appendix F

ANOVA Using Degree of Fulfillment of CBCA criteria as Dependent Variable

ANOVA Using Degree of Fulfillment of CBCA Criteria as Dependent Variable

The ANOVA using degree of fulfillment of CBCA criteria as the dependent variable revealed a significant condition by grade interaction, F (2,141) = 6.02, $p<.005$. Thus, tests for the significance of simple main effects were conducted. There was a simple main effect for experimental condition for Grade 4 subjects, $F(2,136) = 24.47$, $p<.001$. As well, there was a significant simple main effect for experimental condition for Grade 2 subjects, $F(2,136) = 6.59$, $p<.01$.

Follow-up multiple comparisons, using the Tukey method adjusted for unequal n's by the Spjotvoll and Stoline procedure (1973), revealed that for Grade 4 subjects, there was a significant difference in the degree of fulfillment of CBCA criteria by subjects in the LE and LC condition, and by subjects in the HC and LC conditions. LE and HC condition subjects achieved a higher degree of fulfillment of criteria than subjects in the LC condition. There were no significant difference between subjects in the LE and HC conditions. For Grade 2 subjects, there was a significant difference in the degree of fulfillment of criteria by subjects in the HC and LC conditions, with HC condition subjects attaining a higher degree of fulfillment of criteria than LC subjects. There were no significant differences in the degree of fulfillment of criteria by subjects in the LE and HC conditions, or by subjects in the LE and LC conditions.

Further analyses revealed a simple main effect for grade for the LE condition subjects, $F(1,136) = 19.79$, $p<.001$, with Grade 4 LE condition subjects attaining a higher degree of fulfillment of criteria than Grade 2 LE condition subjects. There were no significant grade effects for HC condition subjects.