



8-1-1997

Effects of Expert Statement Validity Assessment Testimony on Lay Evaluations of Children's Statements

Marcus Choi Tye

Follow this and additional works at: <https://commons.und.edu/theses>

[How does access to this work benefit you? Let us know!](#)

Recommended Citation

Tye, Marcus Choi, "Effects of Expert Statement Validity Assessment Testimony on Lay Evaluations of Children's Statements" (1997). *Theses and Dissertations*. 2988.

<https://commons.und.edu/theses/2988>

This Dissertation is brought to you for free and open access by the Theses, Dissertations, and Senior Projects at UND Scholarly Commons. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of UND Scholarly Commons. For more information, please contact und.common@library.und.edu.

EFFECTS OF EXPERT STATEMENT VALIDITY ASSESSMENT TESTIMONY
ON LAY EVALUATIONS OF CHILDREN'S STATEMENTS

by

Marcus Choi Tye

Bachelor of Arts, Princeton University, 1990
Master of Arts, University of North Dakota, 1994

A Dissertation

Submitted to the Graduate Faculty

of the

University of North Dakota

in partial fulfillment of the requirements

for the degree of

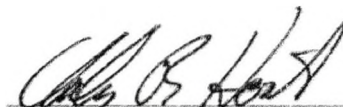
Doctor of Philosophy

Grand Forks, North Dakota

August
1997

997
969

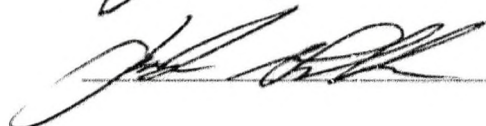
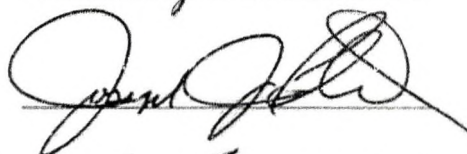
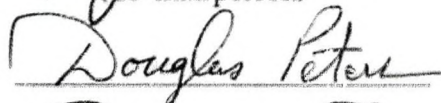
This dissertation, submitted by Marcus Choi Tye in partial fulfillment of the requirements for the Degree of Doctor of Philosophy from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.



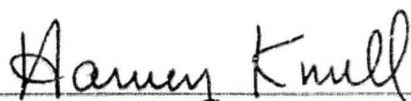
(Co-chairperson)



(Co-chairperson)



This dissertation meets the standards of appearance, conforms to the style and format requirements of the Graduate School of the University of North Dakota, and is hereby approved.



Dean of the Graduate School

PERMISSION

Title Effects of Expert Statement Validity Assessment Testimony on
Lay Evaluations of Children's Statements

Department Psychology

Degree Doctor of Philosophy

In presenting this dissertation in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my dissertation work, or, in his absence, by the chairperson of the department or the dean of the Graduate School. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my dissertation.

Signature



Date

3 June 1996.

TABLE OF CONTENTS

LIST OF FIGURES.....	vii
LIST OF TABLES.....	viii
ACKNOWLEDGMENTS.....	ix
ABSTRACT.....	xi
CHAPTER	
I. INTRODUCTION.....	1
Historical Treatment of Child Allegations of Sexual Abuse.....	2
Defining Abuse: Cross-Cultural Views on Sexual Contact with Children.....	5
Present Day Treatment of Child Allegations of Sexual Abuse.....	10
Prevailing Paradigm: Suggestibility and Memory.....	12
II. STATEMENT VALIDITY ASSESSMENT AND CRITERIA-BASED CONTENT ANALYSIS.....	21
History of Statement Validity Assessment.....	21
Current Form of SVA.....	26
SVA and CBCA Research.....	34
Statements Obtained from the Field.....	35
Statements Obtained from Laboratory Experiments.....	42

Evaluation of Adult Statements.....	50
The Abilities of Untrained Evaluators.....	53
Objectives of the Present Research.....	55
III. STATEMENT MATERIALS	58
Children's Statements.....	58
Expert Testimony	62
IV. METHOD	64
Subjects.....	64
Procedure.....	64
V. RESULTS	69
Preliminary Analyses.....	69
Analyses of Experimental Hypotheses.....	76
Hypothesis 1: Performance of Naive Lay Evaluators	78
Hypothesis 2: Effects of CBCA Testimony on Lay Evaluators	79
Hypothesis 3: Effects of Misleading Testimony on Lay Evaluators	79
Post-Hoc Analyses of Lay Evaluator Performance.....	81
Examination of Secondary Factors and Measures	81
Other Predictors of Lay Evaluation Performance	87
Performance of CBCA.....	90
Presence of Criteria.....	90
Classification Performance	95

Relative Effectiveness of CBCA and Lay Evaluators	99
Hypothesis 4: Comparison of CBCA and Lay Evaluators	99
VI. DISCUSSION.....	101
Shortcomings of Lay Evaluators	101
What SVA and CBCA May Offer	105
What CBCA and SVA Do Offer.....	109
What SVA Did Offer (to Lay Evaluators in the form of Expert Testimony)	111
Implications of Present Findings.....	115
APPENDIX A: SAMPLE STATEMENT.....	121
APPENDIX B: EXPERT TESTIMONY	124
APPENDIX C: EXPERIMENTAL MATERIALS.....	134
APPENDIX D: RESAMPLING STATISTICS.....	146
REFERENCES.....	156

LIST OF FIGURES

Figure

1. Beliefs about Children's Willingness to Lie, Hours of Child Contact	72
2. Effects of Testimony and Medium on Performance	77
3. Testimony and Sex on Which Statement Medium Subjects Preferred	84
4. Effect of Testimony and Subject Sex on Evaluation Confidence	84
5. Effects of Testimony and Statement Reality on Performance	87
6. Subject Evaluations of Testimony Helpfulness and Persuasiveness	88
7. Scores of Criteria by Statement Type: Cumulative Contributions	93
8. Distributions of Mean Sum of Criteria Scores	94
9. Conceptual Continuum of SVA Research	119

LIST OF TABLES

Table

1. National Center for Child Abuse and Neglect Criteria.....	19
2. Three Components of SVA.....	27
3. CBCA Criteria.....	29
4. Validity Checklist.....	31
5. Presence of Criteria in Boychuk/Raskin/Honts Data.....	40
6. Breakdown of Useable Statements by Research Condition.....	62
7. Variables Examined in Each Group of Analyses Conducted	74
8. Summary of Testimony x Medium ANOVA.....	78
9. Testimony x Medium x Subject Sex x Statement Reality MANOVA.....	82
10. Inter-Correlations of Subject Evaluation of Testimony	89
11. Presence of Criteria in Child Statements by Condition.....	91
12. Discriminant Analysis on Expert CBCA Classification.....	96
13. Logistic Regression on Expert CBCA Classification.....	97
14. Comparison Summary of Evaluation Performance.....	99
15. Bootstrap Comparisons of Differential Evaluation Performance	100

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to Charles R. Honts, Ph.D., for his scholarly mentoring and his careful guidance both of this research and my professional development. I am also grateful for Dr. Honts' friendship throughout my graduate school years, and for his continuing his involvement in this project from afar, after having accepted a position at another university.

I would also like to thank my co-chair, Jeffrey E. Holm, Ph.D., for his time and efforts in assuming responsibility for administrative matters. I am grateful too for the efforts of my other committee members, Douglas P. Peters, Ph.D., Joseph J. Plaud, Ph.D., and Jack M. Geller, Ph.D., for their service on my advisory committee, their careful scrutiny of my manuscript, and their valuable input during the development of this research.

I am singularly indebted to the large research team which conducted the experiment on children's willingness to lie and for their allowing me to use their statements in the present research. This was a project designed by Dr. Honts and Dr. Peters and executed by Mary Devitt and Susan Henderson. The children who participated in that study also have my gratitude, as well as the subjects who participated in the present research. My thanks also to Sara Barney and Louise Weller who assisted with my data collection.

I would like to thank Division 41 of the American Psychological Association (Psychology & Law) for their Grants-in-Aid Committee Dissertation Grant. In addition to the financial support of the present research I am also grateful for their recognition of the merit of work in this area. My thanks also to my parents, my grandparents, and the parents of my wife.

I would like to reserve my final thanks for my wife, Una Choi Tye, and to take this opportunity to recognize her insightful assistance with editing, to acknowledge her tireless emotional support, and to commend her unwavering performance as personal "morale officer."

Dedicated to

최운화

ABSTRACT

Research on memory and suggestibility in children continues to demonstrate circumstances under which children may not give accurate accounts of important events. In investigating and treating child sexual abuse, it is important to determine what has happened. Statement Validity Assessment is a witness credibility assessment technique developed by Udo Undeutsch which focuses on the reality of individual statements. It involves a special open-ended investigative interview, Criteria-Based Content Analysis, which searches for the presence of characteristics which indicate that a statement is based on an experienced event, and a Validity Checklist which examines alternate hypotheses. This technique has been widely used in Germany since the 1950s, and since the 1980s has been studied in North America. Many contemporary SVA studies have significant limitations in generalizability to sexual abuse assessment. These methodological problems include using adult statements, using child statements of weak ecological validity, and inadequately trained expert evaluators. In general, lay subjects are only 55%-65% accurate in evaluating the truthfulness of children's statements. This suggests that a role for SVA in the North American legal system might be to improve juror performance with exposure to educational testimony about the general characteristics of a true statement. To assess the potential benefits of such education, 333 undergraduates

were asked to evaluate statements about an emotionally involving mock crime children had witnessed (ages six to ten). Statements were evaluated in videotaped and written form. Prior to evaluation, one-third of subjects received exposure to a videotape of a courtroom-style expert testimony about SVA and CBCA. One-third of subjects received exposure to a counter-CBCA testimony designed to test the effectiveness of the experimental manipulation. The remaining subjects received no testimony exposure. Results indicated that exposure to SVA testimony does not help lay evaluators, who averaged 65% correct. Subjects were less accurate with videotapes than transcripts. These results contrast with expert application of CBCA, which correctly classified 89% of statements. SVA offers powerful tools for the investigation and assessment of sexual abuse. Expert application of CBCA is strongly confirming of truthful statements, which suggests the appropriateness of admitting testimony on the validity of a specific statement.

CHAPTER I

INTRODUCTION

Statement Validity Assessment[†] (SVA) is a technique to determine the truthfulness of a child's allegations of sexual abuse (Undeutsch, 1989). It involves a number of components, including an assessment of motivation and an open-ended interview of the child. The core of this approach is Criteria-Based Content Analysis (CBCA), a technique that focuses on assessing the truthfulness of the statement made by a child during the SVA interview. While commonly used in Germany in sexual abuse cases (Steller & Köhnken, 1989), this approach is relatively new in the United States (Raskin & Esplin, 1992a). Experimental research on SVA has only recently been undertaken (Honts, 1994; Honts, Devitt, Tye, Peters, & Vondergeest, 1995). This dissertation examines the history of children's allegations of sexual abuse and considers how SVA fits with theories of children's memories, determines experimentally whether brief exposure to SVA can improve the ability of lay subjects to evaluate children's statements, and considers these results in the context of the U.S. system of jurisprudence.

[†] In the United States this technique is currently referred to as Statement Validity Assessment by some authors, and Statement Validity Analysis by others. Statement Validity Assessment is used here.

Historical Treatment of Child Allegations of Sexual Abuse

When there is an allegation of child sexual abuse, three immediate concerns come to mind: protecting the child and other children so that future abuse does not occur, treating the psychological trauma resulting from the abuse, and punishing the perpetrator. For many theorists, treating the victim has little to do with punishing the abuser, nor is it necessary to prove that abuse occurred before being able to proceed with treatment. Indeed, the process of proving the abuser's guilt may be exceedingly traumatic to the victim. It may be possible to mediate this trauma by special procedures in courtrooms, such as using videotaped testimony rather than allowing in-court cross-examination (Bulkley, 1992). In any event, if society does choose to punish offenders, the determination of guilt is intimately tied with assessing the credibility of the child's accusation. This question of credibility is addressed in more detail below.

The issue of child sexual abuse has been drawing increasing public attention and professional activity over the last few years (deYoung, 1992). Reports of abuse have been increasing rapidly (Raskin & Yuille, 1989), and ideas about abuse have changed over the years. A few decades ago, the public perception of a perpetrator was an image of an older predatory homosexual male, abducting defenseless children from playgrounds (Newton, 1992). National media attention has broadened publicly available examples of perpetrators to include almost anyone: for example, mothers, priests, day care providers, and

even estranged celebrity fathers engaged in protracted custodial battles. There is also a popular movement whereby psychologically troubled adults, who have no recollections of abuse, can "recover" their allegedly repressed memories of abuse by participating in special therapy groups (Safran, 1993). With this focus on the past and the apparently increasing frequency of parents using false accusations of abuse to strengthen custody cases, there is the danger that real and current abuse will be overlooked.

For abusers to be prosecuted, it is first necessary that people be willing to consider that sexual abuse may occur. The current societal and judicial climate is clearly one in which the possibility of abuse is readily accepted. However, Western society has not always been so open to the suggestion that some adults engage in forced sexual contact with children (Wasserman & Rosenfeld, 1992). In the mid-19th century medical and legal authorities almost universally regarded children's allegations of sexual abuse as false. When presented with physical evidence of vaginal or rectal trauma in children, authorities went so far as to describe the child's injuries as self-inflicted with the intention of harming the reputation of the accused adult (deYoung, 1992).

For a brief period in the 1860s, things started to change. The dean of legal medicine at the University of Paris, Ambroise August Tardieu, conducted court-ordered medical examinations and autopsies of children. He found evidence of sexual abuse that he concluded could not have been self-inflicted. Further, in

children whom he interviewed, he found evidence of emotional pain which he concluded was unrelated to any prospect of material gain. He concluded that the majority of allegations of sexual abuse made by children, including very young children, were true and could be backed up by medical evidence (deYoung, 1992). In the 1880s, the chair of legal medicine at the University of Lyons, Alexandre Lacasagne, investigated with student assistants hundreds of allegations of sexual abuse, and concluded that a child could be systematically abused without any physical trace (deYoung, 1992). Lacasagne had argued for identifying abuse victims on the basis of psychological interviews—on children's testimony alone.

Despite the work of Tardieu, in 1880 Alfred Fournier, an outspoken skeptic of abuse, addressed the French Academy of Medicine urging physicians to regard all uncorroborated allegations of abuse as symptoms of *Pseudologia phantastica*—pathological fantasy (deYoung, 1992). Notwithstanding Lacasagne's subsequent research, this view came to be widely accepted by the medical community, and also affected the emerging field of psychoanalysis.

Freud's initial conceptualization of his patients' neuroses took their self-reports of childhood abuse as fact. At the time he started to introduce his theory to colleagues, the medical community was almost totally convinced that respectable affluent males would never sexually abuse their daughters. Had Freud maintained his theory, he may well have been debunked as a quack.

Perhaps diplomatically, he quickly abandoned his seduction theory and later postulated the Oedipal Complex, and promoted the explanation that children fantasize sexual contact between themselves and parents of the opposite sex. Freud concluded "that he could not distinguish between objective truth and fiction among the repressed ideas that psychoanalysis brought to light" (Wasserman & Rosenfeld, 1992, p. 63). In the absence of confirming evidence, Freud chose a theory safely consistent with the prevailing notion of *Pseudologia phantastica*. In contrast, it is easy for modern commentators to state that at least some of Freud's adult patients may well have been experiencing body memories from real sexual abuse they experienced as children (Wasserman & Rosenfeld, 1992).

Defining Abuse: Cross-Cultural Views on Sexual Contact with Children

An often overlooked cross-cultural perspective on the issue of adult sexual contact with children should be introduced. Certain issues are raised here which are often entirely overlooked in the literature. North Americans have been accused by many of being insensitive to cultural diversity and moral relativism. Accordingly, this section considers some of the issues of cultural relativism concerning childhood sexuality.

Much research on sexual abuse focuses not on treating the victims, nor even on isolating the perpetrators, but on punishing the perpetrators. deYoung (1992) noted that an allegation of sexual abuse has a paradoxical quality: "The

substantiation of its truth in any one case only engenders doubt that truth will be found in any other. Certainty and skepticism are juxtaposed" (p. 256).

In light of this paradox, which reflects the historical disbelief in the existence of abuse, the legal system should try to confine sexual abuse prosecutions to clearly legitimate cases. Thus, what constitutes "legitimate" abuse should be made explicit—and, in law, this is in fact the case. Similarly, one should expect scientific researchers also to clearly define what they mean by abuse. In the literature, however, this is frequently *not* the case. Many research articles and even entire books use the term "sexual abuse" without ever defining what acts they are including. For example, the age of consent varies widely in different countries and an act that warrants clinical concern and criminal penalties in one area, may in another area be regarded as innocuously consensual (Davenport, 1992).

A related issue is the definition of normal behavior and the width of the gray zone that separates healthy play from symptoms of abuse. For example, the National Center on Child Abuse and Neglect (Faller, 1993) considers that sexual arousal in a child is one diagnostic symptom of abuse—a statement in contradiction to some understandings of the healthy development of human sexuality. Kinsey, Pomeroy, and Martin (1948) and Kinsey, Pomeroy, Martin, and Gebhard (1953) found it quite common for pre-adolescent boys and girls to have engaged in erotic play with same-sex partners. Masturbation and orgasmic

responsiveness were also noted in one-third to one-half of pre-adolescent children. Assuming that coercion is absent, how many years must separate participants in such activities before psychologists consider the behavior to be abnormal or dysfunctional? Perhaps by better researching the normal development of sexuality during childhood we can better understand those people who as adults become fixated on children for sexual gratification.

In other species, Bixler (1992) observed that such sexual contacts between sexually mature and sexually immature conspecifics are rare, even among non-human primates. He goes so far as to say that, with the exception of accidental matings, humans are unique in having numerous adults who fruitlessly expend sexual energy on the sexually immature—an activity which can have no reproductive fitness in the Darwinian sense. In way of explanation for a variety of unfit practices (in terms of reproductive fitness), Bixler proposes that, “[V]ery probably, we alone understand the function of the sex act; we conceive of conceiving” (p. 98). He also notes that animals are best fitted to the environments in which they evolved, and that when removed, bizarre results may occur. He suggests that just as animals in a zoo are in an environment for which they did not evolve, modern humans live in industrialized societies that are radically different from the multimale, multifemale hunter-gatherer groupings in which most of our recent evolution occurred. Therefore, he argues it is perhaps not surprising that in humans one sees a wide range of pathology and “unnatural”

behaviors that cannot be explained from a sociobiological view (Bixler, 1992). However, the use of sexual behaviors to establish dominance is readily observed in many primates, and sexually immature members of a troupe may upon maturity become mates of a dominant male (Box, 1984).

Davenport (1992) and Wasserman and Rosenfeld (1992) have detailed adult-child sexual relations in non-industrial cultures. From a cross-cultural perspective, when adult-child sexual relations occur, they are generally part of a developmental stage which is outgrown, and not a fixation of sexual satisfaction for adults. For example, the Sambia people of New Guinea take 7 to 10 year old boys from their mothers, and place them in a special group for 10-15 years. During this time, the boys perform daily fellatio on older males—the elders of the Sambia teach that semen is absolutely vital and should be consumed daily since the creation of biological maleness and the maintenance of masculinity depend upon it. Thus, from middle childhood until puberty, boys perform fellatio on older youths. As they reach puberty, they become the recipients of fellatio performed by younger boys. At marriage (between 16-25 years of age), they may still act as bisexuals. However, as soon as they become fathers, Sambia males become exclusively heterosexual.

Sandfort (1992) took the highly controversial view that not all adult-child sexual contacts in Western society are harmful. He argued that sexual abuse should be clearly defined as occurring only when there is force, or when a power

differential exists. He stated that this is almost always the case when the victims are under the age of 10, but cautions that sexual relations between adults and adolescents should not necessarily be viewed as abusive unless this is the conclusion of the adolescent. He cited research conducted in the Netherlands that shows that many young adolescents (12-14) reported having sexual contact with adults labeled as consensual by the adolescents themselves, and that these children show no poorer psychosocial adjustment than like-aged cohorts who have never had sex or have only had sex with same-aged peers. Sandfort suggested that the absence of harmful sequelae is due to the fact that societal attitudes about human sexuality are somewhat different in the Netherlands than in the United States. Furthermore, he stated that sexual play and experimentation are common in both early childhood and middle childhood. One survey found that as many as 80% of six and seven year olds have been found engaging in sex play with siblings or friends (Kolodny, cited in Rice, 1992). Although our society tends to keep children in same-age peer groups, it may well be natural for children of mixed ages to engage in sex play.

Sandfort also noted that some published studies of sexual abuse label exposure in childhood to erotic magazines abuse. Sandfort argued that victims of traumatic sexual abuse will suffer if the term itself is devalued by being broadly applied. A single encompassing definition of abuse might yield incidence rates so high that the possibility of sexual abuse may again come to be rejected by the

general public (who might not take the time to investigate the definitions used in such studies).

Present Day Treatment of Child Allegations of Sexual Abuse

As Wegener remarked (1989), generally truthful persons may upon occasion lie, and generally untruthful persons may have cause to tell the truth in isolated cases. Thus, in a court of law, the focus should not be on the credibility of a witness but on the credibility of a particular testimony. As discussed above, the current era has seen a widespread professional acceptance that sexual abuse can and does occur. This acceptance includes the spheres of medicine, psychology, and law. However, we do not see the same universal acceptance of the ability of children as a class to testify truthfully. For more than a century, professionals have cast aspersions on the ability of *any* child to deliver a truthful testimony. In the stronger, less politically correct language of 1883, physician Paul Brouardel raised the following doubts about the truthfulness of a child's allegation of sexual abuse:

One often speaks of the candor of children. Nothing is more false. Their imagination likes to invent stories in which they are the hero. The child comforts herself by telling herself fantasies which she knows are false on every point. . . . This child, to whom one ordinarily paid only the most minor attention, finds an audience that is willing to listen to her with a certain solemnity and to take cognizance of the creations of her

imagination. She grows in her own self esteem, she herself becomes a personage, and nothing will get her to admit that she deceived her family and the first people who questioned her. Her lie will be all the more difficult to unmask since the child lies without troubling herself over the improbabilities which one finds in her account. (cited in deYoung, 1992, p. 257)

Wakefield and Underwager (1989)—researchers who also provide treatment for victims and perpetrators, and testify in court cases in the United States—raised this concern about the malleability of children's memories: "Absent strong corroborating legal evidence or third-party witnesses, or without an admission by the accused, admission into evidence by the court of children's statements standing alone must be viewed with great caution" (p. 65). Although the language is couched in a hundred plus years of academic restraint and gender inclusive phrasing, the idea is essentially the same as that expressed by Brouardel: Children as a class cannot be trusted to tell the truth; no unsubstantiated child allegation of abuse is to be believed

Not all contemporary authors condemn the unsupported testimony of all children. Nevertheless, issues of the admissibility of child testimony remain highly controversial. The issue is an important one, because, as Lacasagne noted in the last century, often there is no substantiating physical evidence of abuse. If society wishes to punish perpetrators of abuse, the legal system often must

choose between believing the words of an adult and believing the words of a child. Undeutsch (1989) reiterated that child sexual abuse is frequently nonviolent and non-witnessed. He noted that it is usually committed by someone known to the child, that there may be no emission of semen or other physical evidence of sexual activity, that perpetrators are usually exceptionally careful to ensure that their deviant acts will occur in privacy where no one will notice (certainly where no adult witnesses will notice), and that perpetrators rarely confess unless they are persuaded that others believe the child. Thus, it may well be that in the majority of real cases of sexual abuse one must accept the testimony of the child as legal evidence, or be unable to pursue the matter further in the courts. Central to this decision is the question of the credibility of children's testimonies. Before looking in depth at the testimony-focused approach known as Criteria-Based Content Analysis, this chapter concludes by reviewing the issues and debates in the body of literature on suggestibility, and describing recent trends about the importance of individual differences and developmental models of memory.

Prevailing Paradigm: Suggestibility and Memory

While the central legal issue in many individual cases of abuse is a particular child's testimony, the majority of research into the credibility of children's testimony has focused on the child as witness per se, rather than on the testimony of individual children. As noted above, this paradigm has strong

historical roots. It continues to dominate the present day research, and as yet there is little consensus as to the extent of children's suggestibility (Ceci & Bruck, 1995).

Basic assumptions about the nature of children vary widely. At one end are those who would cast doubt on every child's testimony—who regard children as a whole as “mentally vulnerable” individuals who hold “precariously accessible memories” (Tully & Tam, 1987, p. 187). Goodman and Clarke-Stewart (1992) represent the other end of the continuum—a qualified endorsement that most children will tell the truth, even in the face of leading interviews. In their review of a series of studies on suggestibility, they wrote the following conclusion:

If these results can be generalized to investigations of abuse, they suggest that normal children are unlikely to make up details of sexual acts when nothing abusive happened. They suggest that children will not easily yield to an interviewer's suggestion that something sexual occurred when in fact it did not, especially if non-intimidating interviewers ask questions children can comprehend. However, leading questions in these studies also resulted in a small number of children making errors that could be misinterpreted as suggesting that abuse had occurred. (p. 103)

Goodman and Clarke-Stewart (1992) presented a series of studies which show varying rates of suggestibility in young children. They found that young

children (3- and 4-years-old) are more likely to falsely report abuse when asked questions such as "She touched your bottom, didn't she?" if the interviewer was emotionally neutral. With a friendly interviewer, fewer errors were made. They also found that the presence of anatomically correct dolls did not affect the accuracy rates of both 3-year-old and 5-year-old children, and that some 3-year-old children answering affirmatively to the question "Did he touch your private parts?" were in fact referring to non-private places, such as their ears or arms. Despite the fact that one child falsely affirmed that she had been given a bath, five children falsely agreed that they had been hugged and kissed by the confederate, and two children falsely said "yes" when asked if their picture had been taken in the bathtub, they concluded that young children have high resistance to suggestibility. Other studies presented in the same paper follow a similar pattern: a minority of children make alarming false allegations after an innocuous manipulation. Goodman and Clarke-Stewart (1992) gave this conclusion about the effects of counter-experiential leading interviews on responses to factual questions in a large study:

Overall, the children who heard a biased interpretation inconsistent with what they had observed answered as many of the 17 standard factual questions correctly as the children who heard a neutral or a consistent interpretation. A handful of children (about one fifth [*sic*] of the sample),

however, consistently made errors that were in the direction suggested by the biased interpretation. (p. 101)

On the above, Steller (1992) commented , "Goodman and Clarke-Stewart seem to have ignored their own findings" (p. 108). Goodman's position has changed in recent years. Goodman's team later acknowledged the role of individual differences, and clearly indicated that there are circumstances under which a large proportion of children may be misled (Steward, Bussey, Goodman & Saywitz, 1993).

Many researchers continue to ask the question, "Can children's testimony be contaminated by suggestion?" Ceci, Ross, and Toglia (1987), Doris (1992), Ceci and Bruck (1993), and Lamb, Sternberg, and Esplin (1994) all suggested that the answer is "Perhaps." These authors independently have come to the conclusion that a more productive direction for future work is to elaborate the particular variables that influence accurate recall.

In a series of studies examining this question, Brainerd and Reyna (1995) have shown that children's recall of events is susceptible to what they refer as "gist" intrusions. They demonstrate that children's semantic memory appears to be stored separately from their global perceptions about an experience, and that the latter will interfere with recall. However, this interference does not appear to be an encoding problem, since with appropriate probes that supply cues to the content or verbatim memories, children's accurate recall returns. They find that

this gist interference is much smaller with eight year olds than with five year olds.

The ecological validity of the personal relevance of laboratory research on memory has been questioned by Goodman & Clarke-Stewart (1992), particularly with regards to generalizing to incidents of sexual abuse. Hembrooke and Ceci (1995) presented evidence both affirmative and negative to the question about whether a unique mechanism such as dissociation exists for traumatic memory in children. There are those who claim that older children and even adults can recall memory of abuse events that occurred before language skills were fully acquired (Ceci & Bruck, 1995). It should be noted that Tessler and Nelson (1994) showed that the parental use of language during an experience greatly affected the ability of younger children (three year olds) to have accurate recall. They argued that autobiographical memory development and language are inextricably bound together, and that autobiographical memory would be impossible before language has been acquired.

Recent research seems to be acknowledging the quagmire represented by the advocacy paradigm which seeks to prove that children's recall of serious events is (or is not) impervious to tampering. This new direction acknowledges that a combination of individual differences in development and the characteristics of a situation combine to affect recall. Even more promising is a new focus on the what sort of recall techniques may enhance recall.

Poole (1995a,b) found that after delay, children are more likely to give erroneous information under conditions of free recall. The overall amount of detail—both accurate as well as inaccurate—was greatest under free recall. However, the events the children were in were not personally relevant and she suggested that her subjects may not have attended closely to the situation when it was occurring. An interesting analogy is to be found in the persuasive communications literature. The Elaboration Likelihood Model of communication advanced by psychophysicists and social psychologists Petty and Cacioppo (1986) supports that attention is key to deep processing of information.

The present author suggests that the paradigm that focuses on the various influences on recall is of only limited utility. To almost every question about suggestibility this research paradigm has generated the answer "Maybe, maybe not." For example, there is tremendous controversy about the use of anatomically detailed dolls (Ceci & Bruck, 1995). Under some circumstances and with some children their use enhances recall, and with conditions in the same experiment, it impairs recall (e.g., Gordon et al., 1993; Ornstein, Follmer, & Gordon, 1995). Clinicians making assessments are always working with an assessment of *an individual*, and the accuracy of the assessment rests upon a determination of the validity of an *individual* allegation—a particular *statement*. Thus, statements should be a primary concern for researchers, and it is for that reason that the focus of the present research is on statements.

It is possible that a suggestible child may be abused, and so a research base which speaks to the conditions under which children's testimony can be contaminated is not necessarily diagnostic of whether a particular testimony is accurate or not, even if this research is ultimately able to produce a measure of a particular child's suggestibility. There is in fact such a technique which has as its theoretical underpinnings the hypothesis that recollections that are based on memories of experienced events will differ in quantifiable, detectable ways from recollections that are not based on memories of experienced events. Such a statement-based approach ties in well with the recent work on the development of a narrative memory.

Before at last turning to this promising statement-based assessment approach, it may be helpful to present the prevailing standard for assessing allegations of child sexual abuse in the United States, as advocated by the federal government (see Table 1). This standard relies on three categories of information: a description of the abuse, information about the context of the abuse, and an emotional reaction consistent with the abuse being described.

Some authors (e.g., Saunders, 1991) have argued that therapists should believe an allegation of abuse and treat it as true, and that only legal investigators should be concerned with the factual basis of such a recollection. However, White (1991, 1993) has argued that appropriate treatment depends upon accurate assessment. The need for accurate reality-determining assessment

Table 1: National Center for Child Abuse and Neglect Criteria

The following criteria are given by the National Center for Child Abuse and Neglect (NCCAN) for assessing a child's statements (or behavior) for evidence of sexual abuse (Faller, 1993, p. 54-55):

A Description of the Sexual Abuse:

- sexual knowledge beyond that expected for the child's developmental stage;
- an account consistent with a child's perspective; and
- an explicit description of the sexual acts.

Information About the Context of the Sexual Abuse:

- where it happened;
- when it happened;
- where other people in the family were;
- what the offender might have said to involve the child;
- what the victim and offender were wearing and what clothing was removed;
- the frequency and/or duration of the abuse;
- whether the offender said anything about telling or not telling;
- if so, whom did the child tell and that person's response.

An Emotional Reaction Consistent With the Abuse Being Described:

- reluctance to disclose: characteristic of most children except possibly for very disturbed or very young children;
- embarrassment: a rather mild response often found in disturbed and young children;
- anger: more characteristic of boy victims (but not always evident);
- anxiety: noted frequently in adolescent girls;
- disgust: a typical reaction to oral sex;
- depression: often present in victims who care for the abuser or feel they are responsible;
- fear: typical of cases in which the child has been injured or threatened during the course of the victimization; and
- sexual arousal: another response sometimes found in disturbed and young children

NCCAN also indicates that as many as a third of all accounts may be absent one or more of the categories of criteria.

grows all the more persuasive given that the estimates of false allegations vary widely, from less than 3% in general to as much as 75% in certain circumstances such as custody disputes (Quinn, 1988). The generality of the NCCAN criteria

limit their use in such an assessment—it is a system presented without empirical evidence, whose criteria are not said to be uniquely present in true statements and absent from false ones. Prevailing assessment techniques such as anatomically detailed dolls are often sharply limited, as Lamb (1994) has noted. Medical evidence is often absent even in children who have been subjected to repeated abuse, and behavioral indicators such as sexualized behavior are frequently present in non-abused children and absent in those who are victims of abuse (Lamb, 1994). A reliable assessment technique that could differentiate true and false statements would benefit not only the legal community but also the therapeutic community, and, from all perspectives, the child.

CHAPTER II

STATEMENT VALIDITY ASSESSMENT AND CRITERIA-BASED CONTENT ANALYSIS

Hereafter, when a statement is referred to as "true," it will mean based on memory of experienced events, and when referred to as "false," it will mean not based on memory of experienced events. From a treatment standpoint, there is an important difference between willful deceit and misbelief (fantasy). However, a broader question for assessment (and the only question for a legal proceeding in which someone stands accused of abuse) is whether the child's allegation was essentially based on the memory of an experienced event, and unless specified otherwise false will be used to refer to statements that are either deceitful or the product of fantasy.

History of Statement Validity Assessment

The idea of scientifically analyzing statements can be traced to turn of the century Germany, and the work of William Stern (Undeutsch, 1989). Stern conducted research with adult and child subjects on memory and recall, finding that perfect recall was rare. Stern himself realized the relevance of his research for cases of sexual abuse. The testimony of psychological experts was not, however, readily admissible in German courts. Later, after the second world war, the German judicial system was revamped, and special courts were set up for

offenses committed by or against persons under 21 years of age. In 1954, Udo Undeutsch and a handful of other experts were called by the Supreme Court of the Federal Republic of Germany to consult in a case of adults who had raped a 14 year old girl. The defendants admitted to having had sex with the girl, but denied that the intercourse was forced. Undeutsch questioned the girl and tape recorded his interview. He later replayed the tape in front of the justices, including those sections in which the girl admitted to exaggerating her story. In Undeutsch's words (1989):

The five Justices of the Senate were impressed by the demonstration and convinced themselves that in assessing the truthfulness of the testimony of a child or juvenile witness an expert psychologist conducting an out-of-court examination has "other and better resources" than the persons acting as fact finders within the formal atmosphere of a courtroom trial. (p. 104)

Thereafter, it was common practice for the German courts to summon experts to assist in ascertaining the truth (Köhnken & Steller, 1988; Undeutsch, 1989). The German legal system for trying such cases is not adversarial—there is no jury, and no motivation for a defense and a prosecution to call in "competing" experts. The cases are tried by a panel of judges. The number of judges varies according to the severity of the penalty associated with the alleged offense (Wegener, 1989). It may consist of one judge, one judge and two lay judges, or

three judges and two lay judges. These lay judges hear many similar cases. In this system, it is easy to see how the courts would be willing to consider and even rely upon the assessments of trusted experts. The ethos of the Anglo-American adversarial criminal justice system is not to present neutral unbiased assessments—rather, it is to present two diametrically opposed viewpoints, a forum for them to “duke it out,” and a jury to decide who wins. The jury in the Anglo-American system is not expert, but is almost deliberately naive, lacking in any prior convictions, and the judges are mere referees. A more affluent defendant may conceivably hire more experts and create more doubt in the mind of a jury. In the German system, the people hearing the arguments already possess expertise about the type of case they are considering, and so logically would be better able to evaluate the quality of expert testimony. By extension, the ability of an economically powerful defense to overwhelm the government’s case by drawing in a bewildering plethora of experts would be diminished. On the other hand, the state has more possibility to overwhelm a defendant in Germany than in the U.S., where it is only less affluent defendants who must rely on the patronage of the state for their defense.

Undeutsch has personally been an expert witness in over 1500 cases of child sexual abuse (Boychuk, 1991) and, based on his experience, has argued that he has little respect for the utility of laboratory research in the assessment of credibility. His chief complaint against experimental research is that child sexual

abuse is so vastly different an experience from those that can ethically be created in the laboratory, that there is zero generalizability. Instead, he has developed and refined, through decades of experience, "methods for assessing the truth value of statements" (Undeutsch, 1989, p. 107). Undeutsch has strongly discounted the value of assessing the personality of a witness, or issues related to trustworthiness of witnesses in general. As he said,

To cut a long story short: I have experienced a number of cases in which the witness had all or many of the personality traits that were, according to the traditional rules, considered to be very unfavorable for the chances to obtain a correct and truthful report from this person. . . . Yet in some of these cases during the trial the accused confessed to the crime as charged. (1989, p. 109.)

Undeutsch has focused on assessing the truthfulness of a testimonial statement. He originally referred to his approach as Statement Reality Analysis (SRA). The first English language publication on SRA dates to 1972 (Trankell, 1972). However, journal articles in English on the technique do not appear until the 1980s. SRA included a psychological assessment (gathering background details), a forensic assessment (the interview with the child), and a credibility assessment of the testimony performed according to criteria developed by Undeutsch, Arntzan, Trankell, and Szewczyk (Undeutsch, 1989).

The basic idea of this content analysis is that a child's truthful statement will be different in structure and content from a made up story, and that these differences will be detectable by a blind observer rating a transcript of the child's statement (Boychuk, 1991). Steller (1989) has referred to this as the Undeutsch Hypothesis. Undeutsch proposed that a false statement will likely proceed from beginning to end, with few interruptions, few digressions, and few extraneous details. Children recollecting events that happened to them would, however, tend to jump around in time as they remembered various pieces. Furthermore, there should be a richness of extraneous details to true statements—details not related directly to the accusation. The statement in its entirety would still be logically consistent. The Undeutsch Hypothesis indicates that in children, both willful lies and genuinely believed-in fantasies will be characterized by the same absence of characteristics that will be present only in a statement based on memories of experienced events. Undeutsch does not propose a method for differentiating a lie from a believed-in fantasy.

Steller and Köhnken (Steller, 1989; Steller & Köhnken, 1989) have taken the SRA concepts developed by Undeutsch and others and refined them. The term Statement Reality Analysis has been replaced by Statement Validity Assessment, and the actual assessment of the testimony has come to be formally referred to as Criteria-Based Content Analysis (CBCA). Steller, Raskin, Yuille, and Esplin have further codified CBCA into 19 criteria for analyzing witnesses'

statements. Many of these criteria are the same as used by Undeutsch (Steller, 1989; Raskin & Yuille, 1989; Raskin and Esplin, 1992a), with minor vocabulary changes to make the meaning of terms more self-evident.

Current Form of SVA

Table 2 outlines the three main components of Statement Validity Assessment. SVA begins with an interview designed to maintain objectivity on the part of the interviewer and also maximize the amount of information gathered. Great care is taken in an SVA interview *not* to ask leading questions (and thus to avoid altogether issues of suggestibility), and the interview must be recorded, preferably on videotape. This technique is similar to the type of "special care questioning programme" adopted by the Hong Kong police (Tully & Tam, 1987). This is also consistent with the technique of Vogeltanz and Drabman (1995) who emphasized the importance of reducing contamination by the use of a two-person interview team where one interviewer has no knowledge of the details of a case.

Currently in the U.S., many front-line social service investigators have considerable knowledge about the details of an allegation, prefer to use anatomically correct dolls in their interviews, and are resistant to using open-ended interview techniques because they feel that children will be unwilling or unable to communicate what has happened to them if other investigative techniques are used (Ceci & Bruck, 1995). However, healthy, non-abused

Table 2: Three Components of SVA

-
- Statement Validity Assessment Interview
 - Criteria-Based Content Analysis of statement
 - Validity Checklist
-

children may often play with such dolls in ways that arouse suspicions. As Gabriel wrote, if an examiner is biased and goes on "a fishing expedition" with such dolls, abuse will likely be found (cited in Wakefield and Underwager, 1989, p. 59). This is especially dangerous when the interviewer shuns videotaping, and the manner in which the allegation was elicited cannot be proven. Klitzing (1990) gave examples of cases of known abuse (where medical evidence existed) in which the family, out of shame, told the child to be quiet, and the police could not at first get any testimony from the child. This sort of situation presents difficulty for any interviewing or credibility assessment technique. While SVA presupposes that a child is willing to make a statement, it provides as an integral part of the interview a period of rapport building to facilitate disclosure.

At the conclusion of the open-ended SVA interview, the interviewer may conduct a testing of limits in which the suggestibility of the child is assessed by seeing if the child will agree to saying that something happened counter to the child's previous statement. The statement is then transcribed and analyzed for the presence of criteria. SVA concludes with a validity assessment procedure in

which the child's verbal and cognitive abilities are assessed, as well as the presence of any motivational factors for making a false accusation. This "validity checklist" tests the plausibility of various alternative hypotheses explaining the child's accusation, and is used by the interviewer, in conjunction with the CBCA results, to make a final determination about the truthfulness of the testimony (Raskin & Esplin, 1991). Little research has been performed on the validity checklist, and the literature does not clearly specify what weight the validity checklist should be given in the final SVA decision (Raskin & Esplin, 1991; Raskin & Esplin, 1992a; Horowitz, 1991). Table 3 lists the content criteria and Table 4 gives the validity checklist.

Nevertheless, the core of SVA—and the largest component of the final judgment about the truthfulness of a testimony—is the criteria-based content analysis of the child's statement (Raskin & Esplin, 1991). The generic term "credibility examination" is used by some researchers to refer to an expert's assessment of the psychodynamic motivations for why the child may, or may not, be telling the truth (e.g., Klitzing, 1990). CBCA is a very different technique for assessing credibility. It strives to be a purely objective assessment of the logical structure and content of the child's testimony. In practice, it is somewhat subjective as interpretations must be made. Thus, a short statement would necessarily have difficulty fulfilling a number of criteria because of its brevity. The evaluator must determine whether this indicates a fabrication, or some other

Table 3: CBCA Criteria

The following criteria are given by Raskin and Esplin (1992a, p. 156-157):

Criteria for Analyzing General Characteristics

1. **Logical structure.** Is the statement coherent? Is the content logical? Do the different segments fit together? (Note: Peculiar or unique details or unexpected complications do not diminish logical structure.)
2. **Unstructured production.** Are the descriptions unconstrained? Is the report somewhat unorganized? Are there digressions or spontaneous shifts of focus? Are some elements distributed throughout? (Note: This criterion requires that the account is logically consistent.)
3. **Quantity of details.** Are there specific descriptions of place or time? Are persons, objects, and events specifically described? (Note: Repetitions are not counted.)

Criteria for Analyzing Specific Contents

4. **Contextual embedding.** Are events placed in spatial and temporal context? Is the action connected to other incidental events, such as routine daily occurrences?
5. **Descriptions of interactions.** Are there reports of actions and reactions or conversation? (Note: Verbatim reproduction of conversation is also scored under criterion 6.)
6. **Reproduction of conversation.** Is conversation reported in its original form? (Note: Use of unfamiliar terms or quotes are especially strong indicators, even when attributed to only one participant.)
7. **Unexpected complications during the incident.** Was there an unplanned interruption or an unexpected complication or difficulty?

Criteria for Analyzing Peculiarities of Content

8. **Unusual details.** Are there details of persons, objects, or events that are unusual, yet meaningful in this context? (Note: Unusual details must be realistic.)

9. **Superfluous details.** Are peripheral details reported that are related to the situation but that do not contribute directly to the allegation?
10. **Accurately reported details misunderstood.** Did the child correctly describe an object or event but interpret it incorrectly?
11. **Related external associations.** Is there reference to an event or conversation of a sexual nature that is related in some way to the incident but that did not occur within the incident?
12. **Accounts of subjective mental state.** Did the child describe feelings or thoughts experienced at the time of the incident?
13. **Attribution of perpetrator's mental state.** Is there reference to the alleged perpetrator's feelings or thoughts during the incident?

Criteria for Analyzing Content Related to Motivation

14. **Spontaneous corrections.** Were corrections offered or information added to material previously provided in the statement? (Note: Responses to direct questions do not qualify.)
15. **Admitting lack of memory.** Did the child indicate lack of memory or knowledge of an aspect of the incident?
16. **Raising doubts about one's own testimony.** Did the child express concern that some part of the statement seems incorrect or unbelievable? (Note: Merely asserting that one is telling the truth does not qualify.)
17. **Self-deprecation.** Did the child describe some aspect of his or her behavior related to the incident as being wrong or inappropriate?
18. **Pardoning the perpetrator.** Did the child make excuses for the alleged perpetrator or fail to blame the alleged perpetrator when an opportunity occurred?

Criteria for Analyzing Offense-Specific Elements of Content

19. **Details characteristic of the offense.** Are there elements that are common to this type of offense? (Note: Details contrary to common knowledge are especially strong indicators.)
-

Table 4: Validity Checklist

The following checklist is given by Raskin and Esplin (1991, p. 288):

Psychological Characteristics

1. **Cognitive-emotional limitations.** Are there indications that limited cognitive abilities, unwillingness to discuss the events, or discomfort during the interview interfered with obtaining adequate information from the interview process?
2. **Language and knowledge.** Was the child's use of language and display of knowledge beyond the normal capacity for a child of that age and experience and beyond what the child may have learned from the incident?
3. **Affect during the interview.** Did the child display inappropriate affect during the interview ... as there an absence of affect that would be expected to accompany such a report by this child?
4. **Suggestibility.** Did the child demonstrate susceptibility to suggestion or ask questions during the interview to attempt to obtain clues as to what to say to the interviewer?

Interview Characteristics

5. **Interview procedures.** Was this interview inadequate according to principles and procedures of statement validity assessment? Did the interviewer introduce distractions, fail to establish rapport, inadequately attempt to elicit a free narrative, fail to use open questions and appropriate follow-up questions, or fail to attempt to resolve ambiguities and apparent inconsistencies? Were reasonable alternative hypotheses ignored?
6. **Influence on statement contents.** Was there leading or suggestive questioning, pressure, or coercion in any analyzed interview of the child? Were suggestive techniques or props employed in any interview?

Motivational Factors

7. **Motives for reporting.** Does the child's relationship to the accused or other contextual variables (e.g., living arrangements or relationships among significant others) suggest possible motives for the child to make a false allegation?

8. **Context of disclosures.** Are there questionable elements in the context of the original disclosure or report of the accusations? Are there important inconsistencies in the reports?
9. **Influence by others.** Are there indications that others suggested, coached, pressured, or coerced the child to make a false report?

Investigative Questions

10. **Lack of realism.** Are the described events unrealistic? Are there major elements in the statement that are contrary to the laws of nature?
 11. **Inconsistent statements.** Are there major elements in the statement (not peripheral details) that are inconsistent or contradicted by another statement made by this child or another witness?
 12. **Contradictory evidence.** Are there major elements in the statement that are contradicted by reliable physical evidence or other concrete evidence?
 13. **Characteristics of the offense.** Is the description of the alleged sexual offense lacking in the normal details and general characteristics of this type of offense against a child? Does the description contain important elements or general characteristics that are contrary to what has been established in the professional and investigative literature concerning such offenses?
-

factor, such as a child with poor verbal abilities. CBCA looks at the testimony itself, and not at family dynamics or other reasons for why a false allegation may be made.

The 19 content criteria shown in Table 3 are quite clearly defined and represent the "state of the art" circa 1996. These criteria also represent CBCA as researched throughout the late 1980s and the first half of the 1990s in North America. There is however no single CBCA. As Raskin and Esplin (1992a) point out, the criteria are evolving. The earliest English language text on statement

assessment is Trankell's *Reliability of Evidence* (1972). This volume presents an early form of the present criteria. In addition to semantic and conceptual differences in various authors' presentations of the criteria, no one has accepted a standardized scoring system. Sometimes a two-point scale is used: 0 for absent, 1 for present. Sometimes a three-point scale is used: 0 for absent, 1 for weakly present, and 2 for strongly present. Furthermore, Undeutsch, Raskin, Esplin, and Yuille fail to provide guidelines for what CBCA score is needed to indicate with confidence that a statement is truthful, and what score would cast doubt on the credibility of a statement. There seems to be little consensus as to a clear cut off. Yuille (cited in Joffe, 1992/1994) once required the presence of the first five criteria and any two of the others to classify a statement as truthful. Raskin, Esplin, and Craig (1995) recently suggested using a continuum in which more criteria lend increasing support to the hypothesis that a statement is true. The English language literature does not clearly indicate whether contemporary German investigators use a uniform system or if there are regional or individual differences in CBCA and its application.

Köhnken, Schimossek, Aschermann, and Höfer (1995) have proposed six additional content criteria supportive of a true statement: an issue-related reporting style (versus a long-winded style), the display of insecurities, providing reasons for lack of memory, less use of clichés, fewer repetitions, and

more comments and interpretations regarding the event. These "Köhnken criteria" reportedly increased the classification performance of unaided CBCA.

SVA and CBCA Research

The theory underlying CBCA, that a false statement will differ qualitatively from a true statement, is now referred to as the Undeutsch Hypothesis. Udo Undeutsch himself did not like to do research. He felt that laboratory experiments were of little benefit for developing statement assessment (Undeutsch, 1989):

The growing experience with the testimony in criminal proceedings made it obvious that the laboratory experiences on eyewitness memory bore little resemblance with the situation of a victim-witness of a sex crime and, as a consequence, the results of these experiments have very little bearing on the court room situation. (p. 105)

Undeutsch instead relied on his own extensive use of the technique as evidence of its effectiveness. Of some 1500 testimonies that he had personally assessed, he determined 90% to be truthful, and of these, 95% led to a conviction. In support of the remarkable effectiveness of the technique, he noted that from among all these convictions (approximately 1280), not a single defendant was later found innocent by the discovery of additional, conflicting information. In support further of his confidence in the efficacy of the technique, he stated that over the course of three decades, in many thousands of cases that assessed by

himself and others in Germany and Sweden, there have been no reports of later evidence contradicting the facts contained in a child's testimony—when that testimony had been deemed truthful by statement reality assessment (Steller & Köhnken, 1989). Despite Undeutsch's confidence and this anecdotal record, the widespread use of a technique should be preceded by empirical validation. The growing body of such validating research is described below. For clarity, the literature has been organized to present experiments by research methodology.

Statements Obtained from the Field

There are only a handful of U.S. studies of CBCA that use real statements of sexual abuse. One, conducted by Raskin and Esplin (1992a) yielded very encouraging results. It involved applying CBCA to 40 sexual abuse statements made by children ranging from 3 to 15. Only cases in which the alleged abuser was a friend or relative of the victim were used (on the grounds that there is less motivation to make false allegations involving strangers, therefore these stranger cases would be less likely to contain untrue testimonies [Raskin & Esplin, 1992a]). The efficacy of CBCA was determined by dividing the rated statements into two groups: doubtful statements, and confirmed statements. Confirmed statements were defined as those in which either physical evidence *or* confessions were present (14 cases had both physical evidence and full confessions). The doubtful statements were defined as those in which there was judicial dismissal, a subsequent recantation by the child, *or* a lack of prosecution. In 19 of the 20

doubtful cases, there was no corroborating evidence, in the single case with physical evidence the child later admitted to identifying the wrong perpetrator. Additionally, truthful polygraphs results were obtained in 14 of the 20 “doubtful” cases (that is to say, corroborating the alleged perpetrators’ denials, and indicating that the children’s testimonies were false). Each of the 19 CBCA criteria were assigned a value of 0 if not present, 1 if present, or 2 if strongly present. Scores were summed across all 19 criteria.

In this study, the CBCA scores for the two groups had no overlap whatsoever, thus CBCA was able to discriminate true from false testimonies with 100% accuracy. Confirmed statements received scores between 16 and 34, whereas doubtful statements received scores between 0 and 10 (significantly different, $t(38) = 16.53, p < .001$). Most of the 19 criteria were present in a majority of the confirmed statements, and absent from a majority of the doubtful statements. Furthermore, seven of the 19 criteria were present in every confirmed statement, and seven criteria were absent from all doubtful statements. However, criteria 10, 11, and 16 showed less ability to discriminate confirmed statements from doubtful statements (respectively, these criteria are: accurately reported details misunderstood, related external associations, raising doubts about one’s own testimony). They were present in less than 10% of the confirmed statements. Also, two criteria (1 and 3, logical structure and quantity of details) were present

in more than half of the doubtful statements. No factor analysis or discriminant analysis was conducted.

Wells and Loftus (1992) critiqued this study for circularity. They suggested that the confirmed cases *may* have resulted in plea-bargained confessions and convictions of innocent people because the children who made the allegations gave very convincing testimony. They also suggested that, hypothetically, the "doubtful" cases *may* have been real abuse, but that no convictions occurred because the children in these cases were so unconvincing (which might also account for their low CBCA scores). Raskin and Esplin (1992b) countered that in the doubtful cases there was no medical evidence, and that a majority of the alleged perpetrators passed a polygraph test. However, polygraph test results were not given for the alleged perpetrators in the confirmed group (the ones who were eventually convicted). This debate will always characterize field studies, because the truth can never be known absolutely, and there is always some overlap between the criteria used by the present legal system to determine guilt, and the criteria used by CBCA.

Another field study was conducted by Boychuk (1991), using statements from open-ended interviews that were audiotaped and videotaped in 1987 and 1988. The design was a similar one to Raskin and Esplin's study—testimonies given in cases that had already been resolved were selected for post-hoc CBCA analysis (Raskin and Esplin's field study, described above, was conducted prior

to Boychuk's even though it has a later publication date). Of 97 available transcripts, 75 cases were chosen such that 25 fell into each of three groups: Group A, a confession, clear medical evidence, *and* a conviction; Group B, a confession *and* a conviction (no clear medical evidence); Group C, no confession, a truthful polygraph outcome, no medical evidence, an expert opinion that abuse probably did not occur, *and* judicial dismissal of the case. Thus, one third of the cases were clearly confirmed, one third were strongly confirmed, and one third were clearly doubtful. Whether the middle one third were confirmed or somewhat doubtful depends upon one's point of view. Raskin and Esplin would probably argue they were clearly confirmed, Wells and Loftus might argue that confessions are sometimes made by the innocent in the course of plea-bargaining. Furthermore, abuse may occur with the absence of corroborating medical evidence. Nevertheless, this study was a good field validation of CBCA, particularly because of the strong, clearly designated category requirements for Groups A and C.

Boychuk performed her analysis on the confirmed versus the highly doubtful cases (thus combining Group A and Group B; these two groups were not analyzed separately). Unfortunately, Boychuk only reported results for the statistical significance of the presence or absence of each of the 19 criteria. She did not assign a 0-1-2 weight to the criteria, and did not report criteria-summed scores. She found that 12 of the 19 criteria were present more often in the

confirmed group than in the highly doubtful group ($p < .05$), and that 8 of these 12 were present more often at the $p < .001$ level.

A bootstrap reanalysis of fourteen criteria from this data set conducted by Raskin and Honts (1994) revealed no overlap whatsoever in the sampling distribution of the means, and a near-perfect classification performance. Table 5 summarizes the presence of each criterion in Boychuk's data.

The presence or absence of criteria was affected by many factors, including the age of children, the type of abuse, the recency of abuse, and the number of times the child had been interviewed. She found that the type of abuse significantly affected the presence of criterion 2 and criterion 11 (see Table 3). She also found that the number of times an interview was conducted affected the presence of criteria 1, 2, 3, 5, 9, 14, and 16. She had not expected to find the presence of so many criteria susceptible to change after multiple interviews. Consistent with other literature, Boychuk found that the recency of abuse significantly affected the logical structure and quantity of details, as well as the presence of criteria 2, 4, 11, 14, and 19. Age of child affected the frequency of criterion 15's presence. Two criteria (13 and 17) were never met in the testimony of children less than six years of age.

Brodie (1993) examined the effects of training CBCA and SVA on social workers. She used a set of twelve transcripts of children's statements about abuse (half deemed likely to be true, half deemed likely to be false). She had each of 53

Table 5: Presence of Criteria in Boychuk/Raskin/Honts Data

CRITERIA	Percent of statements with criterion present		
	Confirmed (N=50)		Highly Doubtful (N=25)
<u>General Characteristics</u>			
1 Logical structure	100%	*	68%
2 Unstructured production	100%	*	40%
3 Quantity of details	100%	*	48%
<u>Specific Contents</u>			
4 Contextual embedding	96%	*	44%
5 Descriptions of interactions	66%	*	12%
6 Reproduction of speech	74%	*	20%
7 Unexpected complications	64%	*	8%
8 Unusual details	52%	*	8%
9 Superfluous details	50%	**	24%
10 Details misunderstood	12%		0%
11 External associations	42%	*	0%
12 Subjective experience	64%	*	24%
13 Accused's mental state	10%		4%
14 Spontaneous corrections	86%	*	36%
Mean percent present (of 14):	65.4%	*	24.0%
Mean frequency of criteria:	9.2		3.4
Mean sum of criterion scores:	24.8	*	3.6

Notes: * significantly different, $p < .01$ **significantly different, $p < .05$

county child protective service social workers evaluate two of the transcripts. After seminar training in CBCA, each social worker evaluated another two transcripts and then re-evaluated the first ones according to CBCA. Subjects performed near-chance in both pre-seminar and post-seminar evaluations! The only variable that impacted rater accuracy was the age of the child (they were more accurate with younger children than older children). CBCA training did not improve their performance, and it was found that they did not utilize the statement of the child when reaching their decision—indicating a failing in the subject's use of CBCA. Subject experience, gender of the child, and subject knowledge of abuse did not relate to performance.

The only other U.S. study to use actual abuse statements examined the issue of reliability of scoring. Anson, Golding, and Gully (1993) presented 23 videotapes of investigative interviews of confirmed child abuse victims (ranging in age from 4.1 years to 12.9 years) to four trained CBCA evaluators. Each testimony was evaluated by two of the four evaluators. They found a moderate interrater reliability for CBCA criteria (mean Maxwell's RE coefficient = .49, ranging from -.22 to 1.00 for individual criteria). The authors suggested several possible reasons to account for the poor showing (at odds with reliability in the high .70s and mid .80s reported by Boychuk, 1991): first, the average presence of CBCA criteria was only 41% rather than 70% of the time (substantially less than in other statements). More importantly, evaluators did not receive written

transcripts but only saw videotapes. Perhaps this study demonstrates not that CBCA scoring has low reliability, but the danger of departing from the Undeutsch procedure. Interestingly, Anson et al. also found that age correlated positively with presence of CBCA criteria.

As Wells and Loftus (1992) suggested, CBCA should ideally be assessed in a study evaluating a large number of cases falling on a continuum of clearly confirmed to clearly spurious, including many uncertain cases in the middle. Nevertheless, as Doris (1994) argued, laboratory studies have the benefit of identifying with no uncertainty which statements are really true, and which ones are really false. For this reason they should be an indispensable part of the validation of investigative techniques.

Statements Obtained from Laboratory Experiments

Raskin and Esplin (1992a) criticized studies such as Goodman's for a lack of ecological validity, as they present the child with a situation far removed from the dynamics of sexual abuse. They suggested results from that laboratory research on statement assessment can be generalized to actual sexual abuse investigations if the laboratory scenario possesses the following characteristics: direct involvement of the witness, loss of control, and negative emotional tone. Steller (1989) also emphasized the importance of these three standards for generalizability of statement analysis research. One characteristic of sexual abuse that is hardest to replicate in the laboratory is that the invasively negative event

is often perpetrated by someone known to the child, and often the intensity of the abuse increases incrementally over a considerable period of time (Steller & Köhnken 1989). Unfortunately, strong ecological validity has often been lacking in CBCA laboratory scenarios.

Steller, Wellerhaus, and Wolf (cited in Raskin & Esplin, 1992a) designed a story telling competition with 98 schoolchildren from first through fourth grade. Children were asked to tell two stories: one about an event that had happened to them, and one that they invented. They were limited to topics that had direct involvement, loss of control, and negative emotional tone (such as giving a blood sample, or being attacked by a dog). Children had one week to make up their stories, and were allowed to choose the categories themselves. Prizes were offered to motivate children to tell convincing made-up stories, and parents confirmed which stories were true, and which were false. CBCA was performed on statements gathered in blind interviews. Only the medical-category showed significant differences in CBCA ratings between the true and false stories. Children picked this category the most, and Raskin and Esplin suggest that children making up stories about the other categories (such as being chased by a dog, or getting beaten up by another child) may have seen such events happen to their friends, and so it was little different than if they had actually experienced the event. Also, as Wells and Loftus (1992) noted, the external validity criteria were not really met by Steller et. al's study.

Tye (1994) directly compared the performance of CBCA with that of untrained lay evaluators with a set of 28 statements of known truthfulness from a study by Devitt et al. (1994) on children's willingness to lie. Devitt et al. found that approximately half of children would falsely accuse a researcher of having stolen a book when asked to do so by a parental figure. The statements were about a mock-crime the children had witnessed. Children were asked either by a research assistant or a parent to lie about the theft of a book. In two conditions, children saw the book being taken by the researcher or saw it being taken by their parent. In the latter case, the parents asked the children to falsely accuse the researcher after they came under suspicion. In a final condition, a book mysteriously disappeared without the child noticing and the parent was later accused of theft. The parent then asked the child to make up a story about the researcher taking the book. All three of the Steller/Raskin and Esplin validity demands were met: personal emotional and behavioral involvement, loss of control, and negative emotional tone. The fourth standard, the involvement of a person close to the child, was also met in this scenario. These materials are described in detail in the next chapter.

Tye found that a discriminant analysis on CBCA scores correctly classified 89% of the statements (100% of the false statements, and 75% of the true statements), compared with a near-chance performance of only 56% for lay subjects (49% correct for false statements, and 65% correct for true statements).

The medium of the testimony (videotape of an SVA-style interview or transcript) did not affect subject performance. In this study, lay evaluators were only briefly informed about the experimental scenario which generated the children's statements. Raskin and Esplin (1992b) observed that CBCA is most susceptible to being mistaken when everything a child describes is true, except for the identity of the accused. However, in the study described by Tye, CBCA correctly classified statements as false whether the statement was a pure fabrication about an event the child had not seen, or whether the child was simply substituting the name of a perpetrator. (A reanalysis of the classification was performed for the present research, and is reported on in the Results chapter.)

A follow-up study (Tye, Flonts, & Amato-Henderson, 1994) exposed the evaluators to an experimental scenario in detail so that they might have the same special information about the lab scenario script that the CBCA evaluators had (thus controlling for a possible confound). Subjects were exposed to the scenario complete script by viewing a trial run of the researcher taking the book. They were verbally fully informed about the other two conditions. By being maximally informed about the laboratory scenario generating the statements, lay evaluators were given a "best shot" at maximizing performance. Although overall accuracy went up significantly to 65% correct (62% correct for false statements, and 70% correct for true statements), this was still significantly worse than CBCA. In this follow-up, subjects were significantly more accurate with transcripts than with

videotapes, suggesting that their ability to use the additional information was interfered with when evaluating a videotape. An especially interesting finding from both of these studies was that subject confidence in evaluations was not related to accuracy of evaluations. In fact, not only was the correlation between the two non-significant, it was near-zero!

Despite these mostly promising results, the relative importance of the various criteria has not been agreed upon, nor has a consistent scoring system been adopted. For example, in unpublished research, supplementation was only observed by Littmann and Szewczyk in 8.7% of cases, spontaneous corrections in 20.7% of statements, and the quantity of details was relevant in only 35.8% of the cases they looked at (cited in Steller & Köhnken, 1989). While some criteria appear to be highly frequent and others to be relatively rare, recent research has used relatively small sample sizes. Caution is thus indicated in discarding any criteria. Additionally, no clear definition exists of what constitutes a truthful score, nor how scores should be scaled to account for differences in the quantity of testimony available. Yuille (cited in Joffe, 1992/1994) regarded the first group as necessary but not sufficient. As noted before, exceptionally short testimonies (e.g., from very nonverbal children) logically do not present the opportunity for many criteria to be met, regardless of the truth of the testimony. While CBCA attempts to be objective, whether a statement meets a given criteria or not is partially a subjective judgment of the evaluator. For example, the presence of

inconsistencies might reveal an untruthful statement, or, on the other hand, might be due to poor cognitive abilities or emotional trauma in the child. SVA currently allows only for subjective assessments of the child's verbal and intellectual functioning. The degree to which CBCA can be "fooled" by coached children (or by memories changed by previous suggestive interviews) is also an important issue.

Joffe (1992/1994) conducted an experiment in which lay evaluators were extensively trained in CBCA and before making assessments of children's statements. The goal of the study was to see experimentally test the Undeutsch Hypothesis. Two groups of children (eight year olds and ten year olds) were in one of three conditions. The first was a Live Event (LE), in which children were actively involved in a staged event. The second was Heavily Coached (HE), in which children did not experience the event but were told in detail about, and were specifically told details that would constitute CBCA criteria. The third condition was Lightly Coached, in which children were provided a brief account of the event but had to supply their own details. They were then asked to recall the event in individual interviews, and transcripts of these were analyzed by ten undergraduate evaluators who had received extensive training in CBCA. The scenario involved a child participating in an experiment during which a lamp broke. The experimenter left the room to get a maintenance person, who then came in and fixed the lamp. Next he noticed that a tape recorder was broken, and

he made a point of telling the child that he would take it to be fixed. He had the child help him put the tape recorder into a backpack. Later the experimenter returned and said that the tape recorder contained an important tape, and she needed some help to figure out what which maintenance person had been in the room—this was the pretext for interviewing the child.

Joffe attended to the rigor of the training, noting commentary by Yuille and Undeutsch about the importance of extensive training for adequate use of the technique. Initially evaluators received four two-hour sessions personally conducted by CBCA-expert Yuille. After tests showed that subjects could not reliably identify criteria the four with the highest scores on homework CBCA coding were selected for additional training, and the others were assigned to a minimally-trained control. The highly-trained subjects then received three additional four-hour meetings, with homework, and a training manual until subjects were performing at a high level of accuracy (as evaluated by Yuille by Joffe).

Both highly and weakly trained evaluators had a classification performance of 50% (at chance). The individual criteria ratings of trained evaluators were also analyzed, and as was predicted there were significantly more criteria in the LE than the LC conditions, and no difference between LE and HE. Interestingly, this means that the highly trained evaluators were at least partially effective at identifying the presence of criteria but still did not use this

in their final decisions. For the younger children, there were no differences in criteria across conditions. Perhaps the considerable difference between Joffe and Tye's findings are due to the far greater emotional involvement in the mock crime scenario than in the broken tape recorder scenario. This would suggest that Undeutsch's concerns about the irrelevance of (most) laboratory research are fully justified!

Memon, Cronin, Eaves, and Bull (1993) examined the amount of information obtained from children in a cognitive interview compared with a traditional (structured) interview. The cognitive interview is a technique involves the mental recreation of the original context of an event, the instruction to report everything, recall in a variety of orders, and the description of an event from different perspectives. It also directs the interviewer to begin with rapport building, to be sensitive to the structure of the interview (moving from open-ended to closed questions), and to pay attention to non-verbal responses. It therefore has some similarity to an SVA interview. They found little improvement in information content with the cognitive interview. It should be noted that their event—a vision test in school—failed to meet the Steller/Raskin and Esplin standards of generalizability mentioned above.

A study by Vondergeest (1992) found that expert testimony on SVA did not greatly influence evaluations of the guilt of an accused person in a mock trial video tape created by Margaret Bull-Kovera and Eugene Borgida. This trial was

fabricated and the child's statement was scripted to ambiguous and performed by an actor, which limits the generalizability of Vondergeest's findings.

Evaluation of Adult Statements

Several researchers have looked at the effects of training lay individuals in CBCA. Most of these have had the evaluators examine adult statements. It should be emphasized that the application of CBCA to adult statements is a marked departure from the German use of CBCA. These studies may add insight into CBCA, but any generalization to child statement assessment should be done with extreme caution.

A study by Landry and Brigham (1992) looked at CBCA training in naive subjects and their ability to evaluate statements made by adults. In a 2x2 design, half of the 114 subjects received brief forty-five minute training in CBCA and half did not. Training was conducted in a single session with a group leader discussing each of the criteria, and then illustrating their application to a sample statement. Subjects then evaluated the truthfulness of twelve videotaped statements about a traumatic experience made by adults (half of which were true, and half of which were false). Subjects trained in CBCA performed significantly better than chance and significantly more accurately than those who were not trained in CBCA. Ten of 14 criteria were significant discriminators (as scored by the recently trained undergraduates). Despite these promising results,

they are of uncertain generalizability as subjects were evaluating adult statements.

Ruby (1995) examined racial differences in the presence of criteria in statements. He argued that CBCA was developed on white Europeans and that it might not apply to statements from African Americans. His hypothesis was based on the observation that verbal behavior is learned, and that learned traditions vary across cultures. He found small differences in the overall presence of criteria by race of speaker, and no differences in the overall presence of criteria by the truthfulness of the statement. However, he found that a discriminant analysis of the criteria was able to differentiate true and false statements. He also found that the optimal function depended on the race of the speaker. However, Ruby's technique had several flaws that greatly limits the utility of his findings. Most importantly, he used adult speakers. Secondly, the true and false statements were of greatly varying content as African American subjects were allowed to generate any true or false statement. Statements from adult whites were drawn from a larger archival set to approximate the content of the statements from six African American subjects. Statements varied considerably in emotional intensity (for example, a description of a parent dying of aids, a description of breaking up with a girlfriend). Third, only brief training was provided to the evaluators. Raters were trained for forty-five minutes by an undergraduate research assistant explaining the basis of the technique as they

followed along a written outline. Fourth, the racial difference may have been an artifact of the statements as race was not randomly assigned to statement. In his discussion, Ruby first suggests that his use of transcripts of statements meant that raters could not utilize nonverbal information, and only later points out his nonstandard application of CBCA.

Köhnken, Thurer, and Zoberbier (1994) have examined the use of the cognitive interview with adult subjects. They found that the use of the cognitive interview with adults significantly increased the recall of correct information about a film about blood donation, without increasing the recall of erroneous information. A more conventional structured interview generated less correct recall. However, Köhnken, Schimmesek, Aschermann, and Höfer (1995) found that a cognitive interview increased both accurate recall as well as confabulation. In this study they applied CBCA to a set of statements about witnessing a film of a blood donation episode and found that a discriminant analysis correctly classified 85% of the statements (adult subjects had either seen a film about blood donation, or pretended they saw a film about blood donation). A conservative jackknife resulted in an overall classification accuracy of 73% for CBCA, a rate which is high considering the use of adults, the relatively innocuous subject matter, and the ease with adults in the fabrication condition might have been able to incorporate information from past medical experiences. When including their six additional criteria, discriminant classification improved to 93% (with a

jackknife accuracy of 75%). The following section presents research on the performance of lay adults making evaluations of children's statements.

The Abilities of Untrained Evaluators

Jurors have a propensity to believe that mental health professionals can tell whether a child has been sexually abused or not. In one study, 98% of jurors surveyed thought this was the case, and 82% thought professionals could make this judgment even in the absence of corroborating evidence (Corder & Whiteside, 1988). Brodie (1993) has shown that in fact professionals perform near chance, and fail to utilize CBCA training. However, what of the natural abilities of lay evaluators?

Some of these studies described above contained controls who received no special education, however many of these evaluations were made about adult statements. There are a few relevant studies which consider the abilities of lay evaluators to tell whether children are telling the truth or not (these did not even examine the question of training, and simply looked at "natural" lay evaluation performance).

As noted above, Tye (1994) found lay subjects to be 56% correct in determining whether or not children were telling the truth about a crime they witnessed and were asked by the thief to lie about (the thief was either a stranger or a parent/guardian, this experiment is described in more detail below as the statements obtained from it were used in the present research). Medium (written

or video) did not affect overall subject performance. There were no differences according to the age of the child, although subjects were more likely to rate boys as being truthful (which may have been an artifact of the statements, although the presence CBCA was unaffected by child sex).

In somewhat less emotionally involving scenarios, Vrij and van Wijngaarden (1994) found accuracy rates of approximately 57%, using video clips of children who were asked to convince an interviewer that a drink tasted good even if it did not (two drinks, a pleasant and an unpleasant one, were used; later, the children's actual evaluations of the drinks were recorded to verify the experimental manipulation). Vrij and van Wijngaarden also found similar results in follow-up experiment which had lay evaluators assess video clips of children who were responding to suggestion, or responding free of suggestion, in describing a film they had seen showing a man being angry to a child. Westcott, Davies, and Clifford (1991) had lay subjects evaluate videotapes of children being interviewed about a field trip (some children had taken a field trip to a museum, some had not and were asked to pretend that they had visited a museum). Overall accuracy rates of 59% were found.

All of these studies found that a subject's confidence in an evaluation is unrelated to the subject's accuracy. Across widely different stimulus materials evaluated on two continents, Tye (1994), Vrij and van Wijngaarden (1994), and Westcott, Davies, and Clifford (1991) for the most part found non-significant

correlations between accuracy and confidence. In the few cases in which there was a significant correlation, it was always less than .1, suggesting that lay evaluators are rather poor judges of the truthfulness of children's statements.

Objectives of the Present Research

Joffe (1992/1994) noted that the technique requires extensive training to use well, and both Undeutsch and Yuille have found that evaluators who have had only a few days' of exposure to the technique do not apply it well (cited in Joffe, 1992/1994). As described above, Brodie (1993) also found low utilization of CBCA among U.S. social workers exposed to the technique. Although statement analysis is a psychological specialization in Germany (Wegener, 1989), U.S. courts are reluctant to have an expert testify as to the credibility of a witness or a statement (Ruby, 1995). However, the standard federal rules of evidence will allow expert testimony if it is relevant and deemed by the trial judge to be helpful to a lay jury member (Honts & Perry, 1992). These rules would allow an expert witness to educate a jury of lay persons about the criteria that are thought to *generally* characterize a true statement, and CBCA may most readily be introduced into the U.S. legal system in the form of expert testimony (Honts, 1994). No previous research has looked at the effects of exposure to such expert testimony in CBCA on lay evaluation of children's statements. If such abbreviated exposure to SVA helps lay evaluators, it should be adopted quickly,

given the apparently poor assessment abilities of naive lay evaluators (and the absence of any other assessment techniques).

The present study thus seeks to extend earlier work by Tye (1994) and answer the question of whether exposure to CBCA expert testimony can enhance the ability of lay evaluators to make assessments of children's statements. Other studies have looked at the effects of brief didactic training in CBCA on lay evaluators' performance, but none have looked at the effect on evaluation performance of exposure to the testimony of an expert witness in a form similar to that actually given in court. Further, previous studies have asked subjects to evaluate children's statements of very low emotional saliency, or to evaluate adult statements, or to evaluate an acted statement. The present research is significant in asking lay evaluators to make assessments of actual children's statements about an emotionally involving event.

It might be noted that short-term, passive exposure to CBCA seems to be only minimally effective even for professionals. Yuille and Joffe (cited in Joffe, 1992/1994) have shown that even extensive training in CBCA is insufficient to match the skills of Germany psychologists who study SVA for years before becoming court-appointed experts. However, Landry and Brigham (1992) have shown that relatively brief exposure to the technique can be quite effective. Additionally, passive educational testimony about CBCA has already been introduced into evidence in U.S. sexual abuse trials. This, coupled with the

difficulty of implementing a German-style use of CBCA in U.S. courts, speaks to the pressing importance of the present research which examines the specific effects of exposure to expert testimony (Honts, 1994).

Four hypotheses were formulated for the present study. It was hypothesized that the performance of naive evaluators would be similar to previous findings, and would be found to be slightly better than chance. Secondly, it was hypothesized that exposure to expert-witness testimony on CBCA would improve the performance of lay evaluators, and that this would be proportional to how well subjects attended to the testimony. Third, it was hypothesized that expert testimony per se would not improve performance—specifically that exposure to a misleading testimony that parallels the factual testimony would lower lay subject performance. Fourth, the classification performance of expert application of CBCA was expected to be superior to the performance of both naive lay evaluators and those who had been exposed to testimony.

CHAPTER III

STATEMENT MATERIALS

The child and expert statement materials that were used in the present research are of particularly high ecological validity. Undeutsch's suspicion of laboratory research on credibility assessment has been supported by studies which weakly operationalize the experimental variables. The statements used in this research are described in detail below.

Children's Statements

This experiment used videotaped children's statements that came from a previous study completed at the University of North Dakota by Devitt et al. (1994). This study involved children in a situation in which they had an incentive to make false allegations. Although the children were not themselves directly victims of unpleasant actions, the situation was emotionally charged and highly involving for the child. Two conditions involved making false allegations to protect parent/guardians, in all conditions the child was an active participant in a negative event. Upon debriefing, all children believed that a crime had taken place. Therefore, the ecological validity and applicability to real sexual abuse victim testimony may be as high as it is ethically possible to obtain in the laboratory. Subjects were between the ages of 6 and 10.

The study was explained to parents in full, and their cooperation enlisted in instructing their children that they (the children) were going to take part in some testing of their abilities at the university. After arriving at the psychology building, the child subject and parental figure were escorted into an experimental room by two research assistants (Researcher 1 and Researcher 2). In that room, a student-confederate was "studying for a test." One of the two researchers asked the "student" to leave, letting him or her know that he or she could leave his or her study materials behind, including a textbook needed for the student-confederate's next-day exam. (This was spoken about in front of the child-subject.) Then, Researcher 2 had the child complete a number of tasks—a card sort, drawing a picture, being weighed and measured. During this time, Researcher 1 took the parent aside. While Researcher 1 and the parent were out of the room, one of three situations occurred. Parents knew about the three conditions beforehand, and had agreed to be randomly assigned to any of them.

Situation 1: True Accusation. In this condition, Researcher 2 observed the student-confederate's textbook that had been left on the table, and remarked to the child how nice the book was. Researcher 2 picked up the book, admired it, and then told the child that he or she was going to take the book. Researcher 2 told the child that the theft was to be their secret and that the child should not tell anyone that the book had been taken. Situation 2: Fantasy Accusation. In this condition, the book was removed while the child was not looking. The child and

parental figure were left alone in the room for a few moments. Later, the book was "discovered" to be missing by the returning student-confederate, and a scene occurred in which the student became very agitated and accused the parent of taking the book. The child and parent were again left alone, and this time the parent asked the child to protect him or her by falsely accusing Researcher 2 of stealing the book. Situation 3: Perpetrator Substitution. As in Situation 1, the child actually saw the book being taken, but this time, the parent stole the book during a time when he or she was left alone with the child. The guardian instructed the child that it was "to be their secret," and that the parent would "get in trouble" if the child disclosed what really happened. Then the parent asked the child to falsely accuse Researcher 2.

Later, the child was taken to a "campus police officer" who "had been called" in an adjacent room and interviewed using the SVA technique. These interviews were videotaped. To sum up, in Situation 1, True Accusation, the child was asked to remain silent about a theft. Most children (appx. 85%) in this condition failed to lie on behalf of the research assistant, and made a true accusation. In Situation 2, Fantasy Accusation, the child was asked to make a false accusation to protect the parent (some children may have believed the researcher took the book since a book did disappear, some may not have believed this to be true; in either case the children did not see anyone commit a theft). Approximately two-thirds of the children in this situation did make a false

statement (falsely accusing the research assistant of stealing the book to protect the parent, the others were honest and said they did not know what happened to the book). In Situation 3, Perpetrator Substitution, the child was asked by the parent to falsely accuse the researcher as having taken the book in order to protect the parent, even though the child had seen the parent take the book. About half of the children in this condition falsely accused the research assistant (the others either revealed the truth, that their parent stole the book, or failed to make any accusation). At the end of the study, children were debriefed in full. (They were praised if they told the truth and if they made a false accusation, they were told that it is often difficult to tell the truth, that many children also didn't in the experiment, but that it is best to tell the truth.)

Across all three conditions, this previous study yielded twenty-eight videotaped statements which contained at some point an allegation against the research assistant. Those allegations were either true or false, and thus were usable as source materials for the present study (see Table 6).

For the present study, the source materials consisted of the videotaped interviews of those children's statements in which allegations were made, and printed transcripts made from them. Appendix A contains a sample testimony transcript (the reader is invited to try to determine which condition it came from). Each videotaped interview lasted for approximately 10-12 minutes. Subject names were deleted from the transcripts but left in the videotapes.

Table 6: Breakdown of Useable Statements by Research Condition

TESTIMONY CONDITION	BOYS	GIRLS	TOTAL
True Accusation	4	8	12
Fantasy Accusation (false statement)	4	5	9
Perpetrator Substitution (false statement)	4	3	7
TOTAL	12	16	28

Expert Testimony

The CBCA training took the form of an expert testimony delivered by Charles R. Honts, Ph.D., an experimental psychologist who has studied SVA and CBCA with Undeutsch and Raskin as well as co-authoring several studies on the subject. The videotape lasted for approximately fifty minutes and was modeled after actual testimony he has given in trials. The tape was made in a Virginia courtroom with Dr. Honts on the witness stand, and off-camera an attorney asked questions.

The testimony was in three sections. In the first section, Dr. Honts spoke of the general nature of SVA. In the second section, he addressed each of the criteria and provided examples of them. In the third section, he spoke of research support for CBCA and SVA and the nature of scientific evidence. Dr. Honts'

testimony was prepared in advance and was delivered without the use of a text. A transcript of this preparatory text is to be found in Appendix B (this transcript is therefore not a verbatim copy of the actual testimony).

A counter-training condition was prepared to mislead subjects. The purpose of this was to see if training per se caused subjects to pay more attention, and also to verify any effects of the CBCA training. The counter-training lasted for approximately forty-five minutes and consisted of the same first and third sections. In the second section of the tape Dr. Honts testified that the presence of the criteria signified a statement is false. The criteria were presented in order, as with the true training tape, and changes were only made to the text where necessary for the statement to appear reasonable. (For example, in the misleading tape it was explained that children who suffer abuse are so traumatized by the event that they are unable to recall a great quantity of detail, therefore detail signify a false statement.)

CHAPTER IV

METHOD

Subjects

Undergraduate adults attending beginning and intermediate psychology classes at the University of North Dakota were recruited for participation in a study involving assessing the truthfulness of children's testimonies ($N = 333$, mean age 22 years, 63% female). Subjects received extra credit points in proportion to the length of the research condition they participated in. Subjects were informed that they were eligible to compete for a \$50 prize based on the accuracy of their evaluations. In addition, subjects in the testimony conditions were informed that they were eligible to compete for a second \$50 prize based on how well they remembered the testimony. Prizes were offered both as an incentive for subjects to participate, and to increase attention and motivate subjects to value the experiment as personally relevant.

Procedure

Subjects were recruited to participate in an experiment on assessing the truthfulness of children's statements. They were informed in advance that they would be evaluating videotaped and written statements, and that some subjects might first receive exposure to information that might be helpful in making their

evaluations. Subjects were run in groups according to a randomized mix of testimony conditions and testimonies. At the start of the experiment subjects were informed that they would be evaluating statements about an experiment in which a child witnessed a book being taken. Subjects were told that at some point during a subsequent interview a child would accuse a researcher of taking a book. Subjects were also informed in both verbal and written directions that each child's accusation was either true, and the children had seen the book stolen by the researcher, or it was false. Appendix C contains the consent form, testimony hand outs, and questionnaires.

Subjects in the testimony conditions viewed either the CBCA or the counter-CBCA testimony. After a five minute break they then evaluated four written child statements, followed by four videotaped child statements (each subject thus evaluated a total of eight different statements; randomly, approximately half of subjects evaluated the videotaped statements first). The design was therefore 2x3: statement medium (written or video), by testimony condition (none, CBCA, or counter-CBCA). Pilot runs determined that a total of eight evaluations were possible within a ninety minute period. Accordingly the twenty-eight statements were randomly assigned to seven groups, thus allowing for eight different statements to be evaluated in a given experimental run (four video and four different written). Each of the seven runs were repeated three times (once under each testimony condition).

A resampling power analysis revealed that having each statement (appearing uniquely once in each of the six cells) evaluated by fifteen different people would yield a 91% probability of finding a main effect for testimony condition, a 98% probability for finding a main effect for medium, and a 68% probability of finding an interaction effect, under the assumption that a difference from chance in evaluation accuracy actually existed and was large enough to account for 1% of the variance in performance. An effect of this magnitude would give a correlation of .1 between evaluation decision and reality, and would correspond to an experimental population performance of 55% correct (the null hypothesis population would perform at chance, 50% correct). This was determined to be sufficient power, and accordingly each experimental run was repeated until at least fifteen subjects had participated.

Biographical information was collected from subjects at the beginning of the experiment, and after each statement subjects were asked to make a three-decision forced-choice, selecting from the following: the child was telling the truth about who stole the book, the child was willfully lying, or the child believed what they were saying but the accusation was false. This made it possible in data analysis to examine give evaluators the possibility of identifying fantasy accusations, and yet at the same to allow for collapsing across the last two responses to make a post-hoc true/false dichotomy. The true/false evaluation permitted comparison of lay subject evaluation with the performance

of CBCA. Subjects were also asked to rate their confidence in each decision on a seven-point scale.

Biographical information included whether or not a subject had served on a jury, whether the subject had children, the estimated number of hours of contact with children between the ages of two and twelve (while the subject was at least fourteen years old), and the subject's opinion about whether or not a child would lie about a crime to a police officer (on a scale varying from children always tell the truth to children cannot be trusted, plus a choice for "it depends on the situation" and one for "it depends on the child").

Subjects in testimony conditions also received a handout listing the CBCA criteria. In order to test for the impact of the availability of additional information on recall, approximately one-third of subjects in testimony conditions (randomly selected) received the handout after the testimony videotape ended, the others received the handout before the testimony started and were encouraged to make notes on it during the testimony. All subjects had the handout available to them throughout the statement evaluation period, and it was then collected before a final questionnaire was distributed.

The final questionnaire tested subjects' basic understanding of the testimony material and their evaluation of Dr. Honts' testimony. A free-recall section was provided to see the number of criteria that could be remembered. Subjects in the no testimony condition received an abbreviated final questionnaire. All subjects

were invited to comment on any other strategies they used to evaluate the statements.

All subjects were debriefed by asking them not to discuss the experiment for the remainder of the semester. (The final questionnaire also asked if they had been previously exposed to the research.) Subjects in the counter-CBCA condition were debriefed by being informed that the criteria are actually believed to be present in true statements, and the reason for the deception was explained to them. The testimony conditions ran for between two and half and three hours. The no testimony condition took between one and a half and two hours.

CHAPTER V

RESULTS

Preliminary Analyses

A preliminary sorting of the data was done to eliminate any subjects who had participated more than once, resulting in a discarding of approximately 10 records. This left 333 valid subjects for subsequent analysis, ranging in age from 19 to 60 ($M=22$ years). Of these, 63% identified themselves as female, and 35% identified themselves as male (the remainder failed to provide a response). This represents an average of 15.9 evaluations for each testimony in each of the six cells (three testimony conditions by two statement medium conditions), in accordance with the power analysis presented in the Methods chapter.

The primary response from subjects that forms the focus of the analyses presented below was the forced-choice decision about the essential truthfulness of each child's accusation. These materials are described in detail in the Materials chapter. Briefly, at some point in each statement, the children accused a research assistant of stealing a book. In the True Accusation condition, a researcher had taken a book in front of the child, and had asked the child to keep it a secret. In the Fantasy Accusation condition, a book had vanished, and later the child's parent (or guardian) was accused of the theft. The parent had then asked the

child to say that the researcher had taken the book. In the Perpetrator Substitution condition, the parent had taken the book in front of the child, and then was accused of stealing it. The parent then asked the child to accuse the researcher to protect the parent. In both Perpetrator Substitution and False conditions, the child's accusation about who took the book was false.

Lay evaluators were given three evaluation choices for each statement, as follows: the accusation was true (and the researcher stole the book), the accusation was a lie (and the researcher did not take the book), or the accusation was a fantasy in which the child believed (but still, the researcher did not take the book).

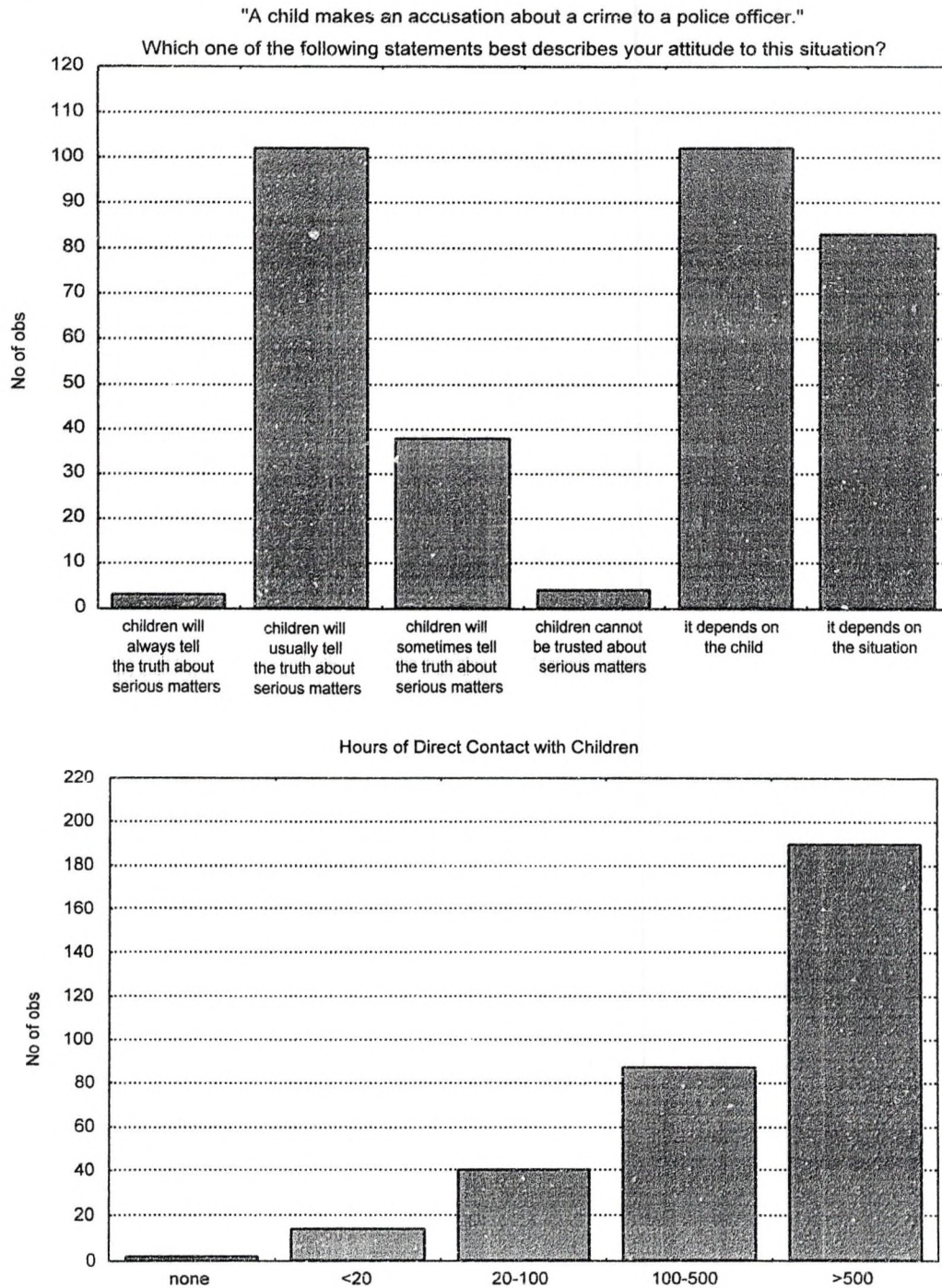
Therefore, each statement has three real possible values (True Accusation, Fantasy Accusation, Perpetrator Substitution) and each lay evaluation has three analogous values. In order to clarify the analysis, where appropriate, the two false possibilities have been collapsed into a single category of "false" (for both statement reality and evaluation choice).

Seven subjects indicated that they had prior experience in assessing the credibility of children, and sixteen had previously served on a jury. Subjects reporting prior experience in credibility assessment had an overall average of 4.6 correct evaluations, compared to 5.0 correct evaluations for the 327 subjects with no prior training. An independent sample *t*-test on prior experience determined this difference to be non-significant, $t(331) = 0.76$, $p = .45$, two-tailed. Sixteen

subjects reported prior jury duty, with these having an overall average of 4.6 correct evaluations, compared to 5.0 correct evaluations for the 317 subjects reporting no jury duty. Again, an independent sample *t*-test on jury duty indicated that this was a non-significant difference, $t(331) = 0.84, p = .40$. Given the small number of cases involved and the fact that these variables did not have a significant independent effect on performance, it was decided to leave these subjects in subsequent analyses on the basis that their presence makes the experimental sample more accurately reflect the experiences of the general population of lay evaluators.

Lay evaluators were asked to indicate whether they thought a child would lie about a crime to a police officer. They were also asked to estimate their total care-giving contact with young children (defined as contact with children ages 2 to 12, after the individual was 14 or older). The responses to these questions are presented in Figure 1. The extent to which these variables predict subject performance is examined below in a multiple regression. There was a near-zero, non-significant correlation between reported hours of child contact and response to children's willingness to lie about a serious matter. Most subjects (91%) reported having no children, the remainder had between one and four children.

The lay evaluator data was coded into two distinct sets. One data set preserved each subject's evaluations (of a subset of the 28 statements) for a logistic regression analysis of factors predicting lay performance. This set is

Figure 1: Beliefs about Children's Willingness to Lie, Hours of Child Contact

referred to as the subject-wise data set. The second data set treated each evaluation as an independent observation, making it possible to perform analyses on the entire set of statements. This second data set therefore contained 2,664 observation/evaluations, and is referred to as the evaluation-wise data set. Matching original subject demographic information and appropriate statement information were coded for each observation to permit more detailed analysis of interaction effects. Table 7 describes the variables that were examined in the following analyses. Some variables appear in either column depending on the analysis, and represent dependent measures that were used post-hoc as grouping factors.

For testimony conditions, two raters examined the free-recall memory questionnaire administered at the end of the experiment. Three variables were coded (the first two were not applicable for the no-testimony condition). The first variable scored was the number of unique responses corresponding approximately to correctly recalled criteria (criteria were the same for both the CBCA and counter-CBCA conditions). The second variable scored was the number of unique responses listed as criteria that met the following two conditions: they were not mentioned in the testimony videotape, and they were clearly counter to the material mentioned in the testimony videotape (commonly occurring examples included "fidgiting" and "shifty eyes"). Responses that were neither clearly criteria nor counter to the testimony were ignored (for example,

Table 7: Variables Examined in Each Group of Analyses Conducted

INDEPENDENT	DEPENDENT
<u>Analyses of Experimental Hypotheses</u>	<u>Analyses of Experimental Hypotheses</u>
Testimony condition	Evaluation (correct/incorrect)
Medium of statement	Criteria correctly recalled
	Whether subjects understood basic premise of testimony
<u>Post-Hoc Analyses</u>	<u>Post-Hoc Analyses</u>
Testimony condition	Evaluation (correct/incorrect)
Medium of statement	Confidence in Evaluation
Actual truth of a statement	Which medium subjects felt was easier to evaluate
Subject sex	Whether subjects thought testimony helped*
	Subject evaluations of expert's competency*
	Opinion about how convincing expert was*
	* no-testimony cases excluded
<u>Regression Analysis</u>	<u>Regression Analysis</u>
Subject grade	Subject performance: recoded as a continuous numerical variable with values of 0-1 (reflecting the percent correct out of eight evaluations)
Subject age	
Subject sex	
Jury duty experience	
Number of children subject has	
Hours of child contact subject has	
Prior training in assessing statements	
Reported belief about whether a child would lie to a police officer about a crime	
Order of statements (videos preceding or following transcripts)	
When subject received testimony handout	
Subject's mean confidence for all eight evaluations	
Which medium subjects felt was easier to evaluate	
Whether subjects thought testimony helped	
Subject evaluations of expert's competency	
Opinion about how convincing expert was	
Whether subjects understood basic premise of testimony	
Number of criteria correctly recalled	
Number of incorrect criteria offered during recall	
Presence of incorrect criteria in open-ended comments	

ambiguous responses such as "repetition"). The third variable, scored for subjects in all three conditions, examined subject responses to an open-ended question soliciting any other strategies subjects may have used in determining which statements were true. This was scored as present if subjects gave one or more common-sense methods which, according to CBCA, are ineffective means of assessing the reality of a statement (e.g., "fidgeting," "shifty eyes," "saying 'um' often"). A subset of records (14%) was scored by both raters, which yielded an inter-rater correlation of .98 for the number of criteria correctly recalled during free-recall, .87 for the number of incorrect responses given as criteria during free-recall, and .94 for the presence or absence of questionable common-sense methods in responses to the open-ended question (all three significant, $p < .05$). This would suggest satisfactory inter-rater reliability, particularly for the most important of the three measures, the total number of criteria accurately recalled.

In order to further test the influence of testimony exposure, a randomly selected subset of one-third of subjects in testimony conditions received the CBCA handout after the videotaped testimony had finished. The other subjects received the handout at the start of the testimony videotape and were encouraged to take notes. All subjects had the handout throughout the statement evaluation period. Subjects who received the handout at the start of the testimony videotape had a mean total number of criteria accurately recalled of

4.85, compared with 3.79 for subjects who received the handout later. This difference approached but did not reach significance, $t(206) = 1.91, p = .057$. This would suggest that subjects were at least superficially attending to the testimony, as the option of taking notes during the testimony increased subsequent recall of criteria.

Analyses of Experimental Hypotheses

Figure 2 shows the main effects of Testimony and Medium on lay evaluator performance, as well as the interaction between these two variables. There was a significant main effect of testimony, $F(2, 2658) = 3.27, p = .038$. Subjects with no testimony exposure had highest proportion of correct evaluations (64.5%), followed by subjects who had been exposed to CBCA testimony (63.4% correct), followed by subjects who had received the counter-CBCA testimony (59.4% correct). The LSD test indicated that only the difference between counter-CBCA and no testimony exposure was significant ($p = .015$). The difference between CBCA and counter-CBCA testimony approached significance ($p = .066$). There was a significant main effect of medium, $F(1, 2658) = 5.65, p = .018$. Subjects performed significantly better with transcripts than with videotapes, with 64.6% correct vs. 60.1% correct. There was no significant interaction effect between statement medium and type of testimony. Table 8 summarizes this ANOVA.

Figure 2: Effects of Testimony and Medium on Performance

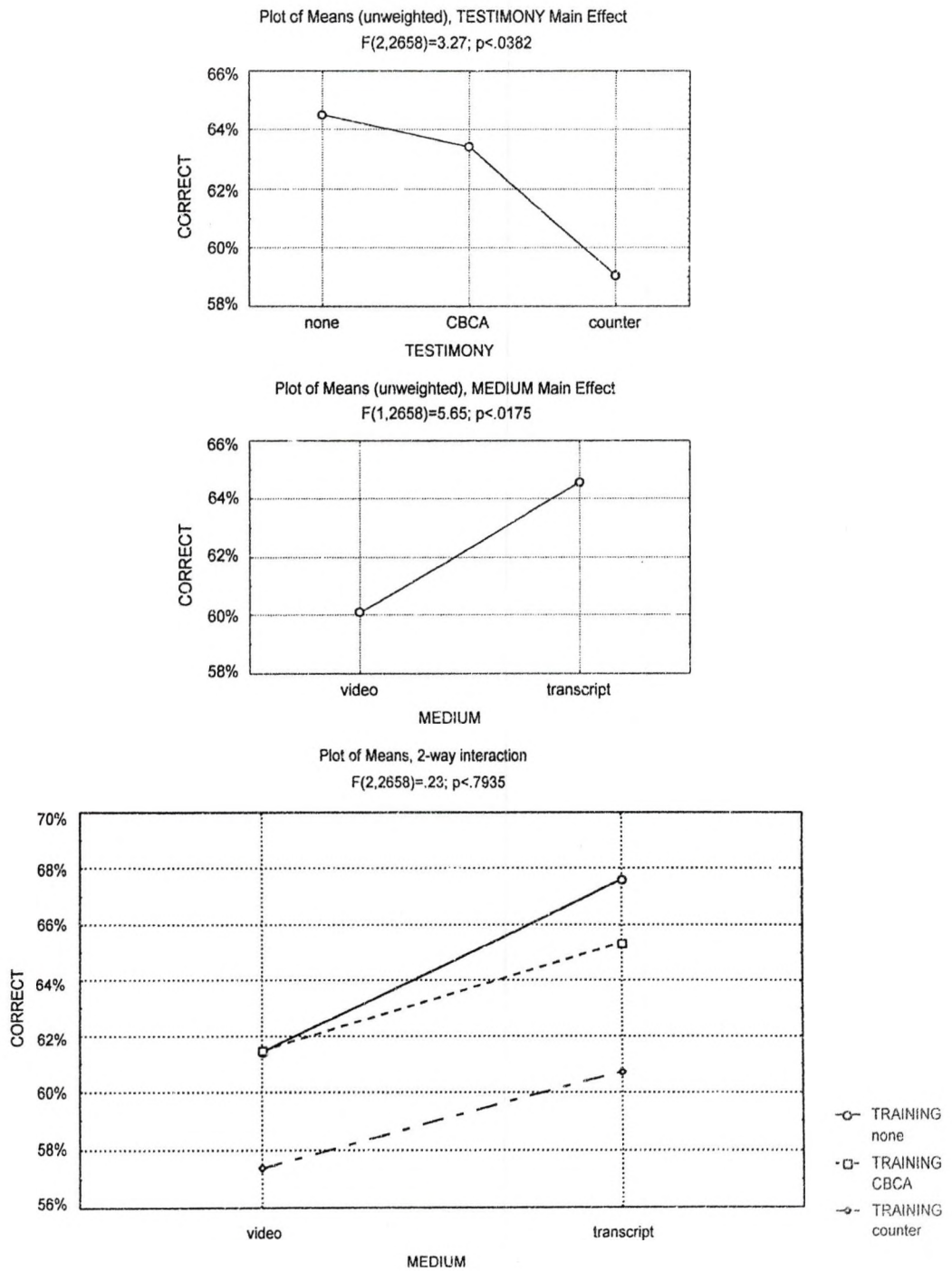


Table 8: Summary of Testimony x Medium ANOVA

	df Effect	MS Effect	df Error	MS Error	F	p-level
Testimony	2	.765170	2658	.234136	3.268059	.038233
Medium	1	1.323979	2658	.234136	5.654743	.017479
Testimony x Medium	2	.054157	2658	.234136	.231304	.793514

Homogeneity of variance was not problematic for the overall design, with a non-significant Cochran's $C = .17$ ($p = .867$). The range of performance across all testimony conditions was 59%-65% correct. Planned comparisons are now presented which compare the effects of testimony exposure against chance performance (50%).

Hypothesis 1: Performance of Naive Lay Evaluators

A resampling analysis was conducted (see Appendix D for a summary of resampling including resampling programs) to directly compare the performance of lay evaluators with no testimony exposure (64.5% correct) against chance. As would be expected with such a large sample size, this reached significance, bootstrap $p < .001$.

Hypothesis 2: Effects of CBCA Testimony on Lay Evaluators

For subjects exposed to CBCA testimony, 63.4% of evaluations were correct, not significantly different from lay evaluators who had no testimony exposure. This performance was also significantly better than chance performance (50%), bootstrap $p < .001$. The correlation between the number of criteria a subject recalled and whether or the subject's evaluation was correct was non-significant and near-zero ($r = -.06$, performed on evaluations in CBCA testimony only), suggesting that subjects who attended to the testimony better did not perform any differently from subjects not attending to the testimony. (A subject-wise analysis of number of criteria recalled and total evaluations correct for CBCA testimony exposure also resulted in a near-zero, non-significant correlation, $r = -.16$.) Subjects were asked to indicate whether the presence of criteria suggests a statement is true or false, 98% of those responding to this question correctly selected true. When subjects received the handout had no effect on accuracy, although as noted above it did significantly effect the recall of criteria. This suggests that subject difficulty in utilizing the testimony was not necessarily a function of memory deficits.

Hypothesis 3: Effects of Misleading Testimony on Lay Evaluators

For subjects exposed to the counter-CBCA testimony, 59.4% of evaluations were correct. This is significantly better than chance, bootstrap $p < .001$, and significantly worse than no testimony (as reported above). This suggests that the

counter-testimony was an effective manipulation that functioned as predicted. Interestingly, the correlation between criteria recalled and performance was also near-zero and non-significant ($r = .02$), suggesting that more close attending to the materials did not increase the deleterious effect of misleading testimony. On the contrary, the testimony apparently only confused subjects—suggested by the fact that, in a final questionnaire, 35% of subjects in the *counter*-CBCA condition indicated that the presence of criteria means a statement is true. The performance of subjects exposed to counter-CBCA testimony was reanalyzed to compare the performance of those who apparently misunderstood the testimony with those who understood it (and were successfully misled). For those who reported that the presence of criteria meant a statement was false (as the misleading testimony indicated), 58.6% of evaluations were correct, compared with 60.2% of evaluations being correct for subjects who reported that the criteria indicate a statement is true. Nevertheless, post-hoc analyses presented below suggest that the counter-CBCA testimony was partially successful in misleading subjects in the expected direction. The difference between the performance of these two was non-significant (bootstrap $p = .662$), and both were significantly better than chance, bootstrap $p < .001$. The correlations between recall and performance were non-significant for both subsets of counter-CBCA subjects.

Post-Hoc Analyses of Lay Evaluator Performance

The children's statements used in the present research were real and not contrived. Given the limited number of these statements and the scenario that generated them, it was not possible to match child characteristics and statement quality across true and false statements. For this reason, in order to be conservative, the planned analyses presented above did not treat statement reality as an independent variable. The post-hoc analyses presented below, however, considers dichotomous statement reality (true/false) as a grouping factor in order to better understand the nature of lay subject performance. In addition, subject sex is also analyzed post-hoc as an independent variable.

Examination of Secondary Factors and Measures

A MANOVA was run with Testimony, Medium, Subject Sex and Statement Reality (true or false) as independent factors. Dependent measures for this analysis were evaluation (correct/incorrect), confidence in evaluation, and which medium subjects indicated they found easiest to evaluate. This analysis examined possible interactions between the most interesting post-hoc grouping variables. Results are presented in Table 9. In addition to the main effects of testimony on number correct discussed above, this revealed a significant interaction effect between subject sex and testimony on subject evaluations of which testimony medium was easier to evaluate. A Scheffé test revealed only one

Table 9: Testimony x Medium x Subject Sex x Statement Reality MANOVA

DESIGN: 4 - way MANOVA, fixed effects
 DEPENDENT: 3 variables: WHICHEASY EVALCNF CORRECT
 BETWEEN: 1-SUBJSEX (2): male female
 2-TESTIMON(3): none CBCA COUNTER
 3-MEDIUM (2): video transcript
 4-ACTUALTF(2): true false
 WITHIN: none

Summary of all Effects; design:
 1-SUBJSEX, 2-TESTIMON, 3-MEDIUM, 4-ACTUALTF

Effect	Wilks' Lambda	Rao's R	df 1	df 2	p-level
1	.979504	17.00536	3	2438	.000000
2	.984660	6.30570	6	4876	.000001
3	.996326	2.99666	3	2438	.029621
4	.989612	8.53057	3	2438	.000012
12	.990242	3.99414	6	4876	.000541
13	.998746	1.02069	3	2438	.382350
23	.999913	.03555	6	4876	.999813
14	.999634	.29716	3	2438	.827475
24	.990594	3.84909	6	4876	.000780
34	.996894	2.53216	3	2438	.055384
123	.999556	.18035	6	4876	.982294
124	.997458	1.03467	6	4876	.400445
134	.998634	.94895	3	2438	.416020
234	.997519	1.00993	6	4876	.416689
1234	.997823	.88603	6	4876	.504038

significant comparison, with more female subjects in the CBCA condition indicating that transcripts were easiest to evaluate than female subjects in the counter-CBCA condition ($p = .035$). This comparison is presented in Figure 3. As can be seen, all subjects thought that videotapes were easier to evaluate than transcripts, and (excepting the contrast already noted) there was a non-significant trend for females to indicate more often than males that videotapes were easiest.

There was no relationship between confidence in evaluations and accuracy of evaluations, with the non-significant correlation of $r = -.03$. Although confidence in evaluations was not related to performance, the ratings of confidence differed according to experimental manipulation. The MANOVA indicated a significant interaction between testimony condition and subject sex on evaluation confidence. As shown in Figure 4, females were slightly less confident than males overall (Scheffé $p < .001$). Testimony tended to increase evaluation confidence compared to no-testimony for both males and females, with the exception of counter-CBCA testimony, which reduced confidence significantly for females but increased it for males (Scheffé $p < .001$). Evaluation confidence was made on a discrete seven point scale (subjects chose one of the following: completely confident, mostly confident, somewhat confident, fifty-fifty, somewhat uncertain, mostly uncertain, completely uncertain; this was then

Figure 3: Testimony and Sex on Which Statement Medium Subjects Preferred

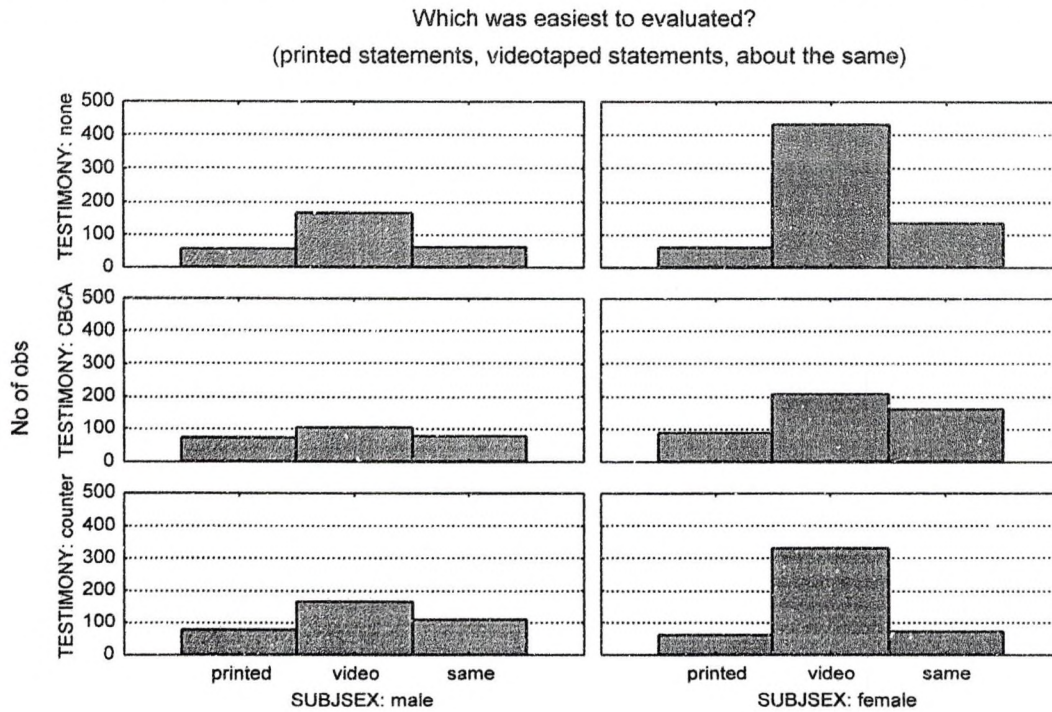
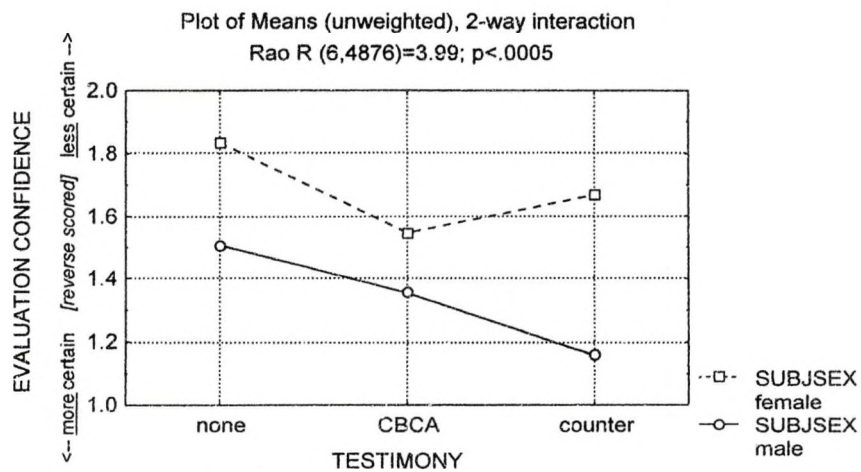


Figure 4: Effect of Testimony and Subject Sex on Evaluation Confidence

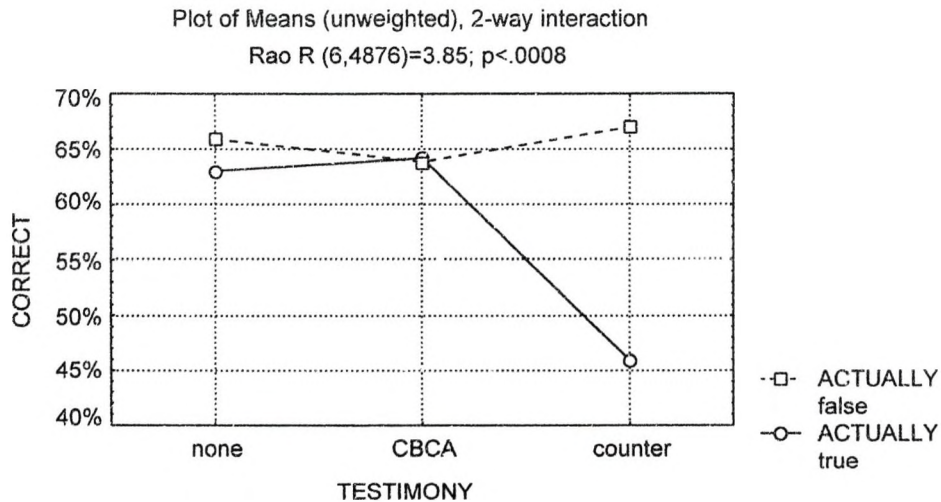


coded from 0-6 with higher numbers indicating increasing uncertainty). The actual difference in confidence is small across all conditions, somewhere in between the judgment "mostly confident" and "somewhat confident." There was also a small relationship between correct recall of criteria and confidence, with subjects who recalled more criteria being slightly more confident ($r = .12, p < .05$ [adjusted to reflect reverse scoring of confidence]).

Follow-up Scheffé comparison tests revealed no significant differences in evaluation accuracy for either subject sex or the subject sex by testimony interaction. Overall subjects were slightly more confident in their evaluations of videotaped statements than in their evaluations of transcripts ($M = 1.46$ vs. $M = 1.57$, respectively, $p = .02$). Subjects were also slightly more confident with their evaluations of true statements compared to false statements ($M = 1.44$ vs. $M = 1.58$, $p = .002$). A number of cell means for evaluation confidence differed significantly for the testimony condition by statement reality interaction. The greatest difference was between subjects who received no testimony exposure being least confident with false statements ($M = 1.77$) and subjects who received counter-CBCA testimony exposure being most confident with true statements ($M = 1.37, p < .001$). In absolute terms these confidence differences are trivial, both rounding to somewhere in between "mostly confident" and "somewhat confident" for decisions about both true and false statements in either transcript or videotaped form under any testimony condition.

The interaction between statement reality and testimony condition had a significant effect on accuracy. Overall, subjects were correct much more often with false statements (65.6%) than with true statements (57.7%), $p < .001$. Furthermore, Scheffé comparison tests revealed that this difference was due entirely to a single combination of testimony condition and statement reality: subjects in the counter-CBCA testimony condition evaluating true statements performed significantly worse (45.9% correct, $p < .001$) than subjects in the other five cells, which were not significantly different from one another (and ranged from 63% to 67% correct). A bootstrap analysis revealed subject performance with true statements in the counter-CBCA condition to be significantly worse than chance, $p = .047$. Figure 5 illustrates the effect of this interaction on performance. This provides evidence that the counter-testimony was effective in misleading subjects.

A second MANOVA was performed to look at the effects of testimony condition and subject sex on subject-wise evaluations of whether or not testimony was helpful, and how competent and convincing the expert was deemed. No-testimony cases were eliminated for this analysis. All three dependent measures were recorded on seven point scales (the full text of the questions appears in Appendix C). The analysis was non-significant, indicating no influences of subject sex or testimony condition on evaluations about the helpfulness and competency of the expert (and also supporting the idea that the

Figure 5: Effects of Testimony and Statement Reality on Performance

counter-CBCA testimony was as convincing as the CBCA testimony). Overall, subjects who received testimony exposure viewed the expert as competent and convincing, and reported that they felt it helped them “a little” in their evaluations. A visual presentation of subject opinion about the expert testimony is presented in Figure 6. The correlation matrix for these three measures is presented in Table 10.

Other Predictors of Lay Evaluation Performance

There was a significant correlation between statement condition (three-level) and lay subject evaluation (three corresponding alternatives), $r = .25$ ($p < .001$). A similar significant correlation ($r = .24$) was observed between dichotomous statement reality (true/false) and the subject evaluation recoded to

Figure 6: Subject Evaluations of Testimony Helpfulness and Persuasiveness

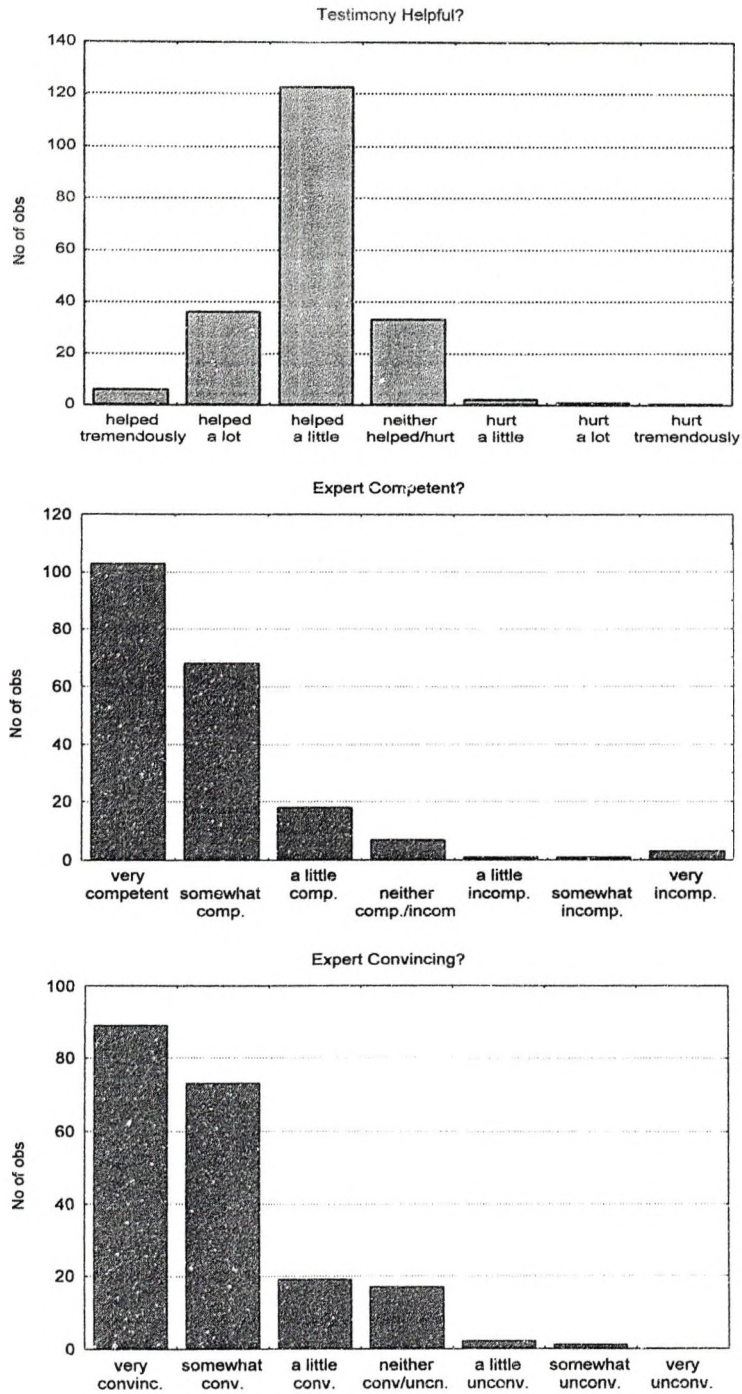


Table 10: Inter-Correlations of Subject Evaluation of Testimony

Marked correlations are significant at $p < .05$
 N=201 (Casewise deletion of missing data)

	Testimony Helpful	Expert Competency	Expert Convincing
Testimony Helpful	1.00	.27	.27
Expert Competency	.27	1.00	.45
Expert Convincing	.27	.45	1.00

group false statements together, indicating that the additional alternatives provided no more information. Separate analyses revealed no other differences between the two false types of statement, or on the three-alternative subject evaluation.

Child sex was examined in a 5-way ANOVA (testimony condition by medium by statement reality by subject sex by child sex). This was not included in the above analyses because child sex was not truly randomly assigned across statements. There was a significant two-way interaction between medium and child sex, $F(1, 2554) = 4.87, p = .027$. The follow-up Scheffé test revealed one significant contrast, with 66.5% of subject evaluations of male children's

videotape statements being correct, compared with 57.2% correct of subject evaluations of male children's transcripts ($p = .017$).

A multiple regression analysis was performed (with each subject serving as a case) in order to see if any subject characteristics would predict subject performance. The total number correct (out of eight possible evaluations) was recoded as a proportion to make a continuous variable to serve as the predictor in the multiple regression. Predictor variables are listed above in Table 7 (confidence was averaged across evaluations to facilitate analysis). Because of the large number of variables, mean substitution was used for missing data (otherwise 47% of the cases would have been eliminated, even though each of these cases had no more than a few missing items). The analysis was not significant, Multiple $R^2 = .057$, $p = .53$. This also supported that testimony exposure did not significantly impact on performance, even considering the effects of correct and incorrect responses during free-recall of criteria.

Performance of CBCA

Presence of Criteria

The panel of expert CBCA evaluators scored the children's transcripts to consensus on a scale of 0 (absent), 1 (present), or 2 (strongly present). Table 11 shows the frequency of finding each criterion in true and false statements, as well as the mean score found for each criterion. Criteria 11 (related external

Table 11: Presence of Criteria in Child Statements by Condition

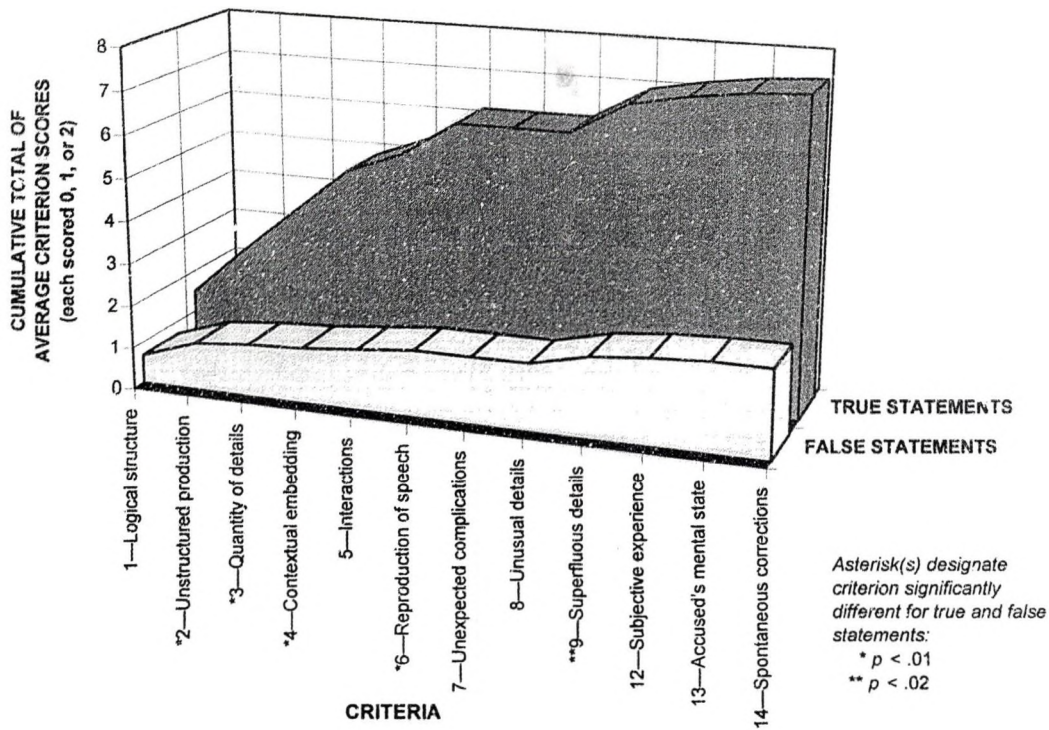
CRITERIA	Percent of statements with criterion present (score of 1 or 2)		Scores for each criterion (0, 1, or 2)			
			True (N=12)		False (N=16)	
	True (N=12)	False (N=16)	M	SD	M	SD
<u>General Characteristics</u>						
1 Logical structure	92%	63%	1.50	0.67	0.69	0.60
2 Unstructured production	92% *	44%	1.17	0.58	0.44	0.51
3 Quantity of details	83% *	13%	1.08	0.67	0.13	0.34
<u>Specific Contents</u>						
4 Contextual embedding	75% *	6%	1.08	0.79	0.06	0.25
5 Interactions	42%	13%	0.50	0.67	0.13	0.34
6 Reproduction of speech	67% *	13%	0.83	0.72	0.13	0.34
7 Unexpected complications	0%	0%	0.00	0.00	0.00	0.00
8 Unusual details	0%	0%	0.00	0.00	0.00	0.00
9 Superfluous details	75% **	31%	0.75	0.45	0.31	0.48
12 Subjective experience	25%	13%	0.25	0.45	0.13	0.34
13 Accused's mental state	17%	6%	0.17	0.39	0.06	0.25
14 Spontaneous corrections	8%	6%	0.08	0.29	0.00	0.00
Mean percent present (of 12):	48%	16%				
Mean number of criteria present:	5.75	2.0				
Mean sum of individual scores:			7.42	3.37	2.06	1.57

Notes: * $p < .01$ ** $p < .02$

associations), 15 (admitting lack of memory), 16 (raising doubt's about one's own testimony), 17 (self-deprecation), 18 (pardoning the accused), and 19 (details characteristic of the offense) were not scored as the panel of evaluators determined these to be particular to statements about sexual abuse, and not salient for the crime scenario. A two-way ANOVA conducted on sum total criteria correct with statement condition (trichotomous) and child sex as independent variables revealed a significant main effect for statement condition, $F(2, 22) = 19.66, p < .001$. Scheffé post-hoc comparisons revealed there to be no significant difference in the presence of criteria in the two types of false statement (for Perpetrator Substitution, $M = 1.0, SD = 0.6$, for Fantasy Accusation, $M = 2.9, SD = 1.6$). True statements were significantly higher in criteria than either type of false statement ($p < .001, M = 7.4, SD = 3.4$). True statements had a median of 8 criteria present and ranged from 0 to 13. False statements had a median of 2 criteria present and ranged from 0 to 6. Figure 7 shows the cumulative contributions of individual criterion scores by statement condition. The main effect of child sex and its interaction with statement reality were not significant.

Figure 8 shows the distribution of the average total score by statement condition, and also presents this information fitted to a normal curve. As can be seen, there is little overlap between the two distributions, with false statements being narrowly distributed around a cumulative sum across criteria of 2, and

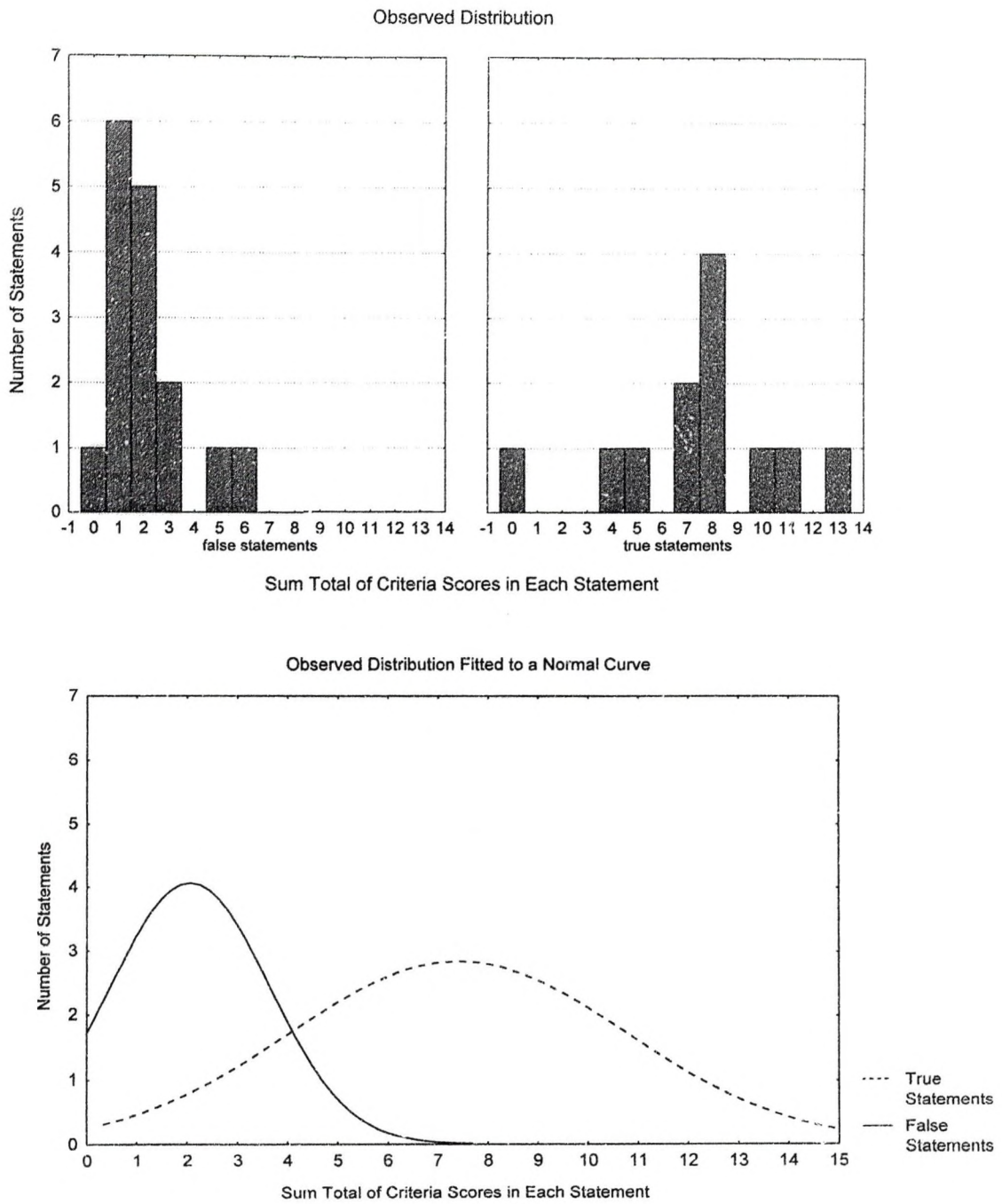
Figure 7: Scores of Criteria by Statement Type: Cumulative Contributions



true statements being somewhat more widely distributed around a cumulative sum of 8. If comparing these sums with other research, note again that each criterion could contribute 0, 1, or 2 points to the sum.

A final analysis considered the effects of length of statement and child age on sum total criteria by examining the inter-correlation matrix for these three variables. Length was measured as the word count for the child's utterances during an interview (after deleting all interviewer questions). There were no

Figure 8: Distributions of Mean Sum of Criteria Scores



significant correlations (largest $r = .23$, smallest $p > .25$), demonstrating that CBCA scoring was unaffected by either the child's age or verbosity. Furthermore, an independent-measures t -test on true and false statements revealed no significant differences in statement length between conditions (false statements $M = 458$ words, $SD = 225$, true statements $M = 471$ words, $SD = 122$; range for all statements was 191 to 843 words).

Classification Performance

A discriminant analysis was performed on the expert CBCA scoring. The sum total of the criteria scores as well as the individual criterion scores were entered into a discriminant analysis using the direct method of entering all predictors at the same time. Criteria 7, 8, and 14 (unexpected complications, unusual details, and spontaneous corrections) were necessarily excluded from the analysis because were absent from all statements. The results of the discriminant analysis are presented in Table 12. Of the 28 statements, 25 were correctly classified (89% correct), with 100% of the 16 false statements being correctly classified, and 75% of the 12 true statements being correctly classified.

A forward step-wise logistic regression was also conducted on the CBCA scores. A logistic regression is indicated for a dichotomous dependent variable. With the present data, the logistic regression however utilized only the total of criteria and its classification performance was slightly worse than the discriminant analysis (86% correct overall). Details are presented in Table 13.

Table 12: Discriminant Analysis on Expert CBCA Classification

Fcn	Eigenvalue	Pct of Variance	Cum Pct	Canonical Corr	After Fcn	Wilks' Lambda	Chisquare	DF	Sig
1*	1.5745	100.00	100.00	.7820	0	.3884	19.859	10	.0306

*Marks the 1 canonical discriminant function remaining in the analysis.

<u>CLASSIFICATION RESULTS:</u>	False Testimonies	—	100% correct (16 of 16)
	True Testimonies	--	75% correct (9 of 12)
	All Testimonies	—	89% correct (25 of 28)

Criterion	Standardized Canonical Discriminant Function Coefficients	Pooled within-groups correlation between discriminant vars. & discriminant function
1 Logical structure	-.04414	.52484
2 Unstructured production	-.06155	.55183
3 Quantity of details	.38636	.77458
4 Contextual embedding	.50050	.76017
5 Descriptions of interactions	.25216	.30122
6 Reproduction of speech	.42517	.54280
9 Superfluous details	-.07797	.38284
12 Subjective experience	-.10217	.13043
13 Accused's mental state	.20969	.13471
14 Spontaneous corrections	.47121	.18164

Table 13: Logistic Regression on Expert CBCA Classification

Method: Forward Stepwise (COND)							
Variable	B	S.E.	Wald	df	Sig	R	Exp(B)
CTOTAL	.7290	.2457	8.8064	1	.0030	.4219	2.0731
Constant	-3.4369	1.1807	8.4741	1	.0036		

CLASSIFICATION RESULTS:

False Testimonies	—	88% correct	(14 of 16)
True Testimonies	—	83% correct	(10 of 12)
All Testimonies	—	86% correct	(24 of 28)

VARIABLES IN THE EQUATION								
Variable	Block 0				Block 1 (after removing CTOTAL)			
	Score	df	Sig	R	Score	df	Sig	R
C1	8.4702	1	.0036	.4113	.8622	1	.3531	.0000
C2	9.0741	1	.0026	.4301	.1714	1	.6789	.0000
C3	13.6015	1	.0002	.5508	.0217	1	.8828	.0000
C4	13.3389	1	.0003	.5445	.6481	1	.4208	.0000
C5	3.5000	1	.0614	.1980	.3026	1	.5823	.0000
C6	8.8728	1	.0029	.4239	.7730	1	.3793	.0000
C9	5.2500	1	.0219	.2915	.2912	1	.5895	.0000
C12	.7304	1	.3927	.0000	.0115	1	.9147	.0000
C13	.7778	1	.3778	.0000	.1415	1	.7068	.0000
C14	1.3827	1	.2396	.0000	.1007	1	.7509	.0000
CTOTAL	15.3561	1	.0001	.5910				

Both a discriminant analysis and a logistic regression may be said to have optimized their classification based on the observed data. A more conservative test of the population performance of CBCA required running a jackknife. The jackknife procedure involves running the analysis with a case deleted, and then using the resulting function to classify this case. Once this has been done for all cases, one has a more robust estimate of the lower bounds of the classification performance. The results of a jackknife on both the discriminant analysis and the logistic regression are presented in Table 14, which also summarizes the initial classification analyses, and the performance of lay evaluators reported above. A resampling analysis revealed that the overall performance of the expert CBCA evaluations (including the jackknifed performances) were better than chance, highest bootstrap $p < .03$ (lowest $p < .001$). The performance for false statements of all expert CBCA evaluations was similarly significantly above chance. Of the evaluations of true statements, only the logistic regression on expert application of CBCA was significantly above chance (bootstrap $p < .02$).

All ways of examining the CBCA classification outcomes yielded significant correlations between classification and reality. The correlation between classification decision rendered by the initial discriminant analysis and statement reality accounted for 63% of the variance in this criterion ($r = .79$, $p < .001$). The least well fitting jackknife accounted for 17% of the variance in the same criterion ($r = .41$, $p = .031$).

Table 14: Comparison Summary of Evaluation Performance

EVALUATION	STATEMENTS			
	Percent variance accounted for	Percent Correct		
		All	True	False
Lay Evaluators, No Testimony Exposure	8%*	65%*	63%*	66%*
Lay Evaluators, CBCA Testimony	7%*	63%*	64%*	63%*
Lay Evaluators, Counter-CBCA Testimony	2%*	59%*	47%*	68%*
Expert CBCA, Discriminant Analysis	62%*	89%*	75%	100%*
Expert CBCA, Jackknifed Disc. Analysis	17%*	71%*	58%	81%*
Expert CBCA, Logistic Regression	50%*	86%*	83%*	88%*
Expert CBCA, Jackknifed Log. Regression	31%*	79%*	67%	88%*

* Cell performance significantly different from chance, $p < .05$.

Relative Effectiveness of CBCA and Lay Evaluators

Hypothesis 4: Comparison of CBCA and Lay Evaluators

Table 15 summarizes the performance of both Expert and Lay applications of CBCA. Lay evaluators did not differ significantly in their performance across the three testimony conditions. Expert-applied CBCA outperformed lay evaluators in all testimony conditions. A CBCA bootstrap analysis was performed to determine whether this higher evaluation accuracy was statistically

Table 15: Bootstrap Comparisons of Differential Evaluation Performance

LEGEND: > signifies significant difference, $p < .05$
 = signifies no significant difference

Lay No Testimony > Lay CBCA Testimony = Lay Counter-CBCA

Discriminant Anal. = Logistic Regression > Lay No Testimony > Lay CBCA Testimony = Lay Counter-CBCA

Discriminant Analysis = Logistic Regression = Jackknife Logistic Regression = Jackknife Discriminant Analysis

Jackknife Logistic Regression = Lay No Testimony = Lay CBCA Testimony > Lay Counter-CBCA

Jackknife Discriminant Analysis = Lay No Testimony = Lay CBCA Testimony = Lay Counter-CBCA

significant (the ratios contrasted in the bootstrap analysis are given at the end of Appendix D). Table 15 displays the contrasts between the differential evaluation performances found in this research. As can be seen, both the discriminant analysis and the logistic regression significantly outperform all lay evaluators (bootstrap $p < .05$). In a separate contrast just of lay evaluators, no-testimony subjects perform significantly better than CBCA-testimony subjects (suggesting that resampling is more powerful than conventional statistics). While both jackknifed expert applications of CBCA accounted for a substantially higher portion of the variance than lay evaluators, this was only significantly better than the lay counter-CBCA testimony evaluation performance. Having a single set of experts provide a definitive application of CBCA limited to absolute number of observations and, with small N , statistical significance is not reached as readily.

CHAPTER VI

DISCUSSION

The present findings suggest that expert testimony in Statement Validity Assessment (SVA) and its credibility testing approach known as Criteria-Based Content Analysis (CBCA) is of little utility to lay evaluators. However, they also indicate that expert application of CBCA is substantially more accurate than lay evaluators. These findings are examined by first considering the performance of lay evaluators, and then by reporting on what SVA can offer to persons who must make credibility assessments with children. Possible reasons for the ineffectiveness of SVA testimony are explored. Lastly, implications for the international SVA literature are discussed with a concluding focus on implications for the U.S. legal system.

Shortcomings of Lay Evaluators

Lay evaluators were found to have three shortcomings. First, they did not evaluate statements with high accuracy. Second, they were unable to monitor their own performance ability. Third, they reported beliefs about children's willingness to lie that are inconsistent with the literature and would tend further reduce their ability effectively evaluate a given statement.

Subjects in the present study were asked whether or not a child might lie about a crime to a police officer. They were able to endorse one of the following statements: children will always/usually/sometimes tell the truth about serious matters, it depends on the child, and it depends on the situation. A majority of subjects endorsed the statement "children will usually tell the truth about serious matters" or the statement "it depends on the child." In contrast, the literature on children's willingness to lie (e.g., Devitt et al. 1994) suggests that "children will sometimes tell the truth about serious matters" is an accurate response. The literature on children's suggestibility and the nature of children's memory (e.g., as reviewed by Ceci & Bruck, 1995) indicates that "it depends on the situation" is most accurate. Only these two responses are congruent with the scientific evidence on children about whether a child might lie about a crime to a police officer. However, less than 40% of subjects in the present study endorsed one of those two responses. These findings suggest that lay people hold the belief that children will generally tell the truth and/or they will discount situational factors. This conclusion is also consistent with the literature. Although the present findings did not investigate the attribution of dispositions such as "honest" or "dishonest" to children, such attributions are suggested by their responses in the present research. This would be lead to an especially devastating outcome for a generally dishonest child who is making a truthful accusation of abuse, and

equally devastating outcome for an innocent person falsely accused by a generally trustworthy child.

As Corder and Whiteside (1988) demonstrated, actual jurors also hold a belief that experts can tell whether or not a child has been sexually abused with near perfect accuracy. However, as research by Brodie (1993) and Ceci and Bruck (1993, 1995) has suggested, experts may well be *inaccurate*. Clearly more research on expert accuracy needs to be done, but guild interests of practitioners conspire against open participation in such research. Furthermore, it would require considerable funding to conduct a meaningful study as a large number of raters and statements would need to be used to achieve generalizability. Police departments, however, may be more open to seeking out techniques for improving their interviewing and investigative performance (Joffe, 1992/1994).

Clearly, then, the ordinary folk who sit on juries in North America have beliefs about children that are unsupported by the research. The legal system, particularly in the U.S., however invests in these people the task of evaluating the testimony—it is for the jury to evaluate the evidence and determine the truth. As Lamb (1994) has noted, sexual abuse generally occurs in private with no witnesses and more often than not there is no penetrative damage or medical evidence to corroborate a truthful accusation. Thus, it rests upon the evaluation performance of the lay subject for the courts to correctly convict those

perpetrators who have been accused, and to free the few innocents who have had allegations made against them.

However, research suggests the accuracy of lay evaluation of children's statements is at best weak, slightly but not greatly above chance. In the present study, the average performance of lay evaluators with no testimony exposure was 64.5% correct. This is one of the highest findings in the literature. Tye (1994) found average performance of only 56%. Vrij and van Wijngaarden (1994) and Westcott, Davies, and Clifford (1991) found accuracy rates also in the mid-50% to 60% range. Others (Joffe, 1992; Ruby, 1995) have also found similar results. Across a variety of statement types and evaluation conditions in several western countries, lay evaluators appear to consistently perform about 5%-15% better than chance. It is particularly important to note that the performance of child subjects is sometimes worse on statements about a serious, emotionally involving event than with statements about something as innocuous as whether a child had a fruit flavored drink.

Even more important is that subjects appear to be unaware of this. This research found subjects being "somewhat" to "mostly" confident of their evaluations, despite the fact that a performance of 65% correct would hardly be considered something to be mostly confident about. More subjects indicated that videotapes were easier to evaluate, however subjects actually performed slightly better with transcripts. This suggests again that lay evaluators are poor self-

monitors of their abilities. Many lay subjects mentioned that they attended to body language and looked for "shifty eyes." Given that these are absent in transcripts and present in videotapes, it would appear that the preferential techniques of lay evaluators are also the least effective.

Further, there was a near-zero correlation between the accuracy evaluations and confidence in evaluations, suggesting again that subjects are relatively unaware of their own performance ability. Imagine the situation on a jury where a subject who is overly confident may argue his or her opinion and persuade others, even though evidently lay evaluators have no innate awareness of how accurate their evaluations are. This near zero relationship is consistent with that found by other authors. In sum, the evidence clearly shows that subjects have room to improve their evaluation performance, both through education in performance-boosting techniques and perhaps through persuasion to abandon pre-existing but ineffective strategies. How much can SVA help them?

What SVA and CBCA May Offer

The North American research on children's ability to give reliable testimony has focused on such global factors as the suggestibility of children and the extent to which interview techniques such as the use of anatomically detailed dolls may mislead children. It has focused also on children's memory for traumatic events and on the decay of memory for both dramatic and trivial

events over time. The focus of this research can be summed up in three words: memory, suggestibility, developmental factors. The author suggests that this represents a paradigm that considers *children* (emphasis on the plural). The author further suggests that the best answer such research can offer about the truthfulness of a *particular* statement from a *particular* child is "Maybe. Maybe not." However, courts do not evaluate children, they evaluate evidence (or at least are supposed to), and the evidence in question in an abuse case is a child's *statement*. German research has focused precisely on statement reality analysis—a paradigm that generates conclusions about particular statements, of direct benefit to the investigator seeking to find out what has happened, especially as appropriate legal and clinical interventions should rest on what has happened.

SVA has been widely used for decades in Germany. In offering a statement-focused paradigm it avoids the "advocacy sand trap" into which the author feels he has fallen when reading some of the literature on suggestibility. Obviously, children vary in suggestibility! But equally obviously, a statement at hand is either true or false. Experimental studies of CBCA show that with careful application by experts well trained in the technique, it is possible to obtain rates of 85%-100%. The classification performance of CBCA as found in the present study depended a great deal on the kind of statistical approach used to determine it. The present study showed accuracy rates in the 80% range, with

statistically optimized classification being at 89% and statistically most conservative classification being at 71%.

It has been argued that CBCA may be most easily "tricked" in the case of perpetrator substitution, because all of the criteria would be present if the child told things exactly as they had happened except for naming a different actor. However, in the mock crime scenario statements used here, Perpetrator Substitution statements had fewer criteria than Fantasy Accusation statements. Even though children had seen their parents steal a book, and heard a script identical to the True Accusations, these children were the least likely to give elaborate, criteria-rich accounts. It is as if, to be sure of not accidentally making a mistake, they completely avoided the memory of the event and fabricated a total substitution.

If this holds true for sexual abuse statements, it could be very helpful for front line caseworkers. In some studies a majority of allegations of abuse are not confirmed (Faller, 1993). This can occur even if abuse *has* occurred, such as when there is insufficient external corroborating evidence. A statement-analysis approach to investigation would help to ensure that it is rare for a child who has been abused to come to the attention of child protective services and to be turned away. The use of SVA would also make it possible to quickly identify possible cases of perpetrator substitution, such as when there is a weak statement that is incongruously accompanied by other evidence that point towards abuse.

Such a wide range of performance is in part due to the lack of definitive cut-off point for CBCA scoring. To a certain extent there is no solution for this problem. Undeutsch (1989) has criticized laboratory research because any ethically created statement of known validity will be substantively different from actual sexual abuse statements. Steller (1989) has cautioned that to be a meaningful test of CBCA, a laboratory scenario generating statements must possess three characteristics: negative emotional tone, loss of control, and direct involvement of the witness. Raskin and Esplin (1992a) added a fourth characteristic: the person behaving negatively towards the child should be someone known and trusted. The Devitt et al. (1994) statements are perhaps the closest that we can come in the laboratory in terms of satisfying these standards--but clearly even having a parent steal a book and asking his or her child to lie is benign compared to sexual abuse.

The discrimination performance of CBCA appears to weaken as these three (or four) "scenario saliency criteria" are diminished. For example, Joffe (1992) found that CBCA performed poorly but used statements about having witnessed a tape recorder malfunction. Steller, Wellerhaus, and Wolf (cited in Raskin & Esplin, 1992a) also found little difference in CBCA scores when children were making up stories about a being attacked by a dog or donating blood. It may be helpful to think of the present statements as at a midpoint between abuse statements and these. Notwithstanding problems of circular

classification, studies that use real confirmed and highly doubtful sexual abuse statements such as Boychuk (1991) and Raskin and Esplin (1992a) have found almost perfect classification performance. If this continuum indeed exists, then the present classification performance is exceptionally high for non-abuse statements. It is argued that future CBCA research should firmly avoid any statements of lesser scenario saliency than used in the present study.

What CBCA and SVA Do Offer

While there is much that SVA may offer, one may well ask what *does* it offer, today, that has been proven? Indeed, Wells and Loftus (1992) have asked such a question and critiqued SVA as a technique that still needs to be experimentally verified. In a general-audience book titled *Jeopardy in the Courtroom*, Ceci and Bruck (1995) had this to say about statement validity assessment:

There is some evidence that professionals can detect outright lies when children are interviewed immediately after lying about an act to protect a loved one (Honts, 1994), but much more work remains to be done before we can be confident about this finding. However, on the basis of the following evidence, it appears that that trained adults cannot reliably detect false reports from children who are exposed to repeated suggestive techniques over long periods of time. (p. 280)

Ceci and Bruck continued by describing a series of experiments using expert evaluators such as researchers, judges, and mental health professionals. Ceci and Bruck made no more than the passing reference to SVA that is quoted above (not even naming the technique), despite the fact that it has had four decades of application in tens of thousands of cases in Germany. Their oversight is not unique, however, as the technique remains little known outside of Germany. Perhaps it is because of Undeutsch's failure to publish prolifically. Nevertheless, science is an international endeavor, and perhaps the reader may excuse this author for not being entirely forgiving of the fact that two of the most prominent North American child witness researchers failed to devote at least a chapter to statement validity assessment, in a book they subtitled *A Scientific Analysis of Children's Testimony*.

As noted earlier, CBCA itself continues to evolve. For example, Köhnken, Schimossek, Aschermann, and Höfer (1995) have proposed a supplementary set of reality analysis criteria which may boost the accuracy of the technique. When studying research on the accuracy, validity, and reliability of CBCA it is important to identify what form of the technique the authors are using. It is also *essential* to consider the type of statements being used in experiments. An increasing proportion of research on CBCA applies the technique in novel ways which are of limited generalizability to child sexual abuse statements. For example, recent research has applied the technique to adult statements or to

children's statements generated in scenarios of low emotional saliency. Similarly, it would be inappropriate to draw conclusions about the reliability of scoring by examining the scoring of briefly trained evaluators, or even by examining the scoring of expert evaluators on statements of low emotional saliency. Recent reviews of statement analysis (e.g., Bekerian and Dennett, 1995) overlook these considerations.

The laboratory evidence that has been reviewed in this dissertation suggests that CBCA *can* discriminate true and false statements, providing that the statements are of sufficient saliency and that the evaluators are specifically trained in CBCA. Indeed, in all the experiments where it has been applied to child sexual abuse statements, it has been perfect in its classification performance.

It is especially important to note that it has been well established that social workers, psychologists, and lay evaluators are all quite poor judges of the truthfulness of children's statements (e.g., Brodie, 1993; Ceci & Bruck, 1995). Thus, there is clearly a need for a performance improving technique. In addition to being such a technique, SVA also has much to offer in terms of its interview system, which may maximize recall while minimizing distortion.

What SVA Did Offer (to Lay Evaluators in the form of Expert Testimony)

As noted above, lay evaluators performed better than chance but not enormously so. Given that CBCA performed much better with the same

statements, one may well ask why exposure to expert testimony about SVA did not make a difference to their performance. It appears that subjects attended at least peripherally to the expert testimony, as they were able at the end of the experiment to recall at least a few of the criteria. Further, subjects who had a handout accompanying the testimony were able to recall about one more criterion by the experiment's end. The fact that a counter-CBCA testimony decreased performance also suggests that subjects were attending to the materials. In both conditions the overwhelming majority of subjects perceived the expert to be either "somewhat" or "very" convincing and competent.

One way of understanding the lack of positive effects for CBCA testimony exposure is by examining the effects of central and peripheral communications, as described in Petty and Cacioppo's (1986) Elaboration Likelihood Model of persuasive communication. Briefly, this cognitive model suggests that there are two routes to attitude change, central and peripheral. Messages that are processed centrally will produce the strongest and most durable change, providing that there is motivation to process the message, available attention, and also the absence of counter arguments (which can cause a centrally processed message to be rejected). Otherwise messages are influenced by peripheral factors such as source credibility, attractiveness, and so forth. Peripheral attitude change is weaker and more fleeting than attitude change due to central communications.

Although the present study is not about persuasion, perhaps expert testimony in the form used in the present study is somewhere between training and persuasive communication. Here, the persuasion is not that of a prosecution or defense witness' opinion, but about the general characteristics of a child's statement—about persuading a lay evaluator (or juror) to abandon whatever techniques they would otherwise use (the ones that cause lay evaluators to perform only a little above chance).

Suppose that people have a set of strategies for evaluating the truthfulness of statements, and that, furthermore, these strategies are not very effective. However, they are closely held, like attitudes. The Elaboration Likelihood Model would predict change to occur only with deep central processing. Accordingly, it would not be surprising for a testimony from an expert who was viewed as competent (a peripheral factor) to have little influence on the underlying attitudes (assessment techniques) unless there was strong motivation to attend to the message. Although subjects had possible financial rewards for their performance, the saliency of the experiment might have been weak for them. While a real juror might benefit more from exposure to expert testimony on SVA, there is also the possibility that the juror might simply continue to rely on his or her pre-existing and performance-lowering strategies, such trying to assess the shiftiness of the witnesses' eyes. (Perhaps there really is a relationship between

lying and eye movements. No matter, since lay evaluators seem to be unable to detect it.)

An alternative explanation for the absence of any beneficial influence of exposure to SVA expert testimony is subject motivation. Perhaps the prize potential was not sufficiently large, or perhaps subjects were not sufficiently impressed by the ability of SVA to improve their performance. Landry and Brigham (1992) found a positive influence of brief training in SVA on lay subject evaluations. Further research would need to be done to determine if the passive nature of exposure to expert testimony inherently diminishes its benefits to lay subjects, or if some motivational factors accounted for the present weak findings.

If motivational factors led to the failure to utilize the SVA testimony, then the present findings are exceptionally important for prosecutors or defense attorneys wanting to introduce expert testimony on SVA in actual sexual abuse cases. For example, Seltzer, Lopes, and Venuti (1990) have noted that jurors fail to utilize expert testimony about the reliability of eyewitness identification. Heeding Undeutsch's warning about weak ecological validity, we may wish to conduct future experiments on SVA expert testimony with circumstances that more closely approximate jury trials, and with real statements from children about events of significant emotional saliency. It would also be important to insist on materials that involve statements of known truthfulness, generated

under conditions that provided child subjects compelling reason to fabricate or tell the truth.

Implications of Present Findings

Lay evaluators do not appear to have been helped by exposure to expert testimony on CBCA. Given that their performance is generally poor, and that CBCA's is better, the present findings do not necessarily argue against the introduction of expert testimony on CBCA in jury trials of sexual abuse cases. Of course, they also do not make a case in favor of this introduction either. The present findings suggest that exposure to expert testimony on SVA is at worst neutral. Manipulations which increased the recall of criteria in lay evaluators did not increase their accuracy, suggesting that the failure to benefit from expert testimony is not a function of memory. As noted above, two alternative explanations are that subjects were unable to apply the criteria, or were unwilling to apply them. Perhaps the heightened saliency and motivation of real jury duty may help. Clearly the present findings show that lay evaluators need assistance, as their native abilities, while better than chance, are not that good. We should not abandon the idea of helping to improve the abilities of lay evaluators with expert testimony.

However, the present findings argue most strongly for the introduction of SVA in courtroom testimony in the form of expert *interpretation* of a child's statement. In just the same manner that a DNA expert presents both educational

information on what DNA tests mean, and data on the application of those tests to a particular sample of blood, an expert testifying on SVA would be able to inform the jury about SVA (as was done with the present testimony materials), and then apply the technique to a particular child's statement. The author suggests that this would not be removing the role of trier-of-fact from the juror, and would be no different than a physician's testimony about the possible causes of genital trauma.

Given that lay evaluators do not perform terribly well, and especially that they appear to unaware of when they are performing accurately and when they are not, it is imperative to begin introducing SVA testimony into U.S. courts of law. Preferentially child protective service workers would adopt SVA interview techniques, as used in Germany for decades, and the expert would be called upon by the prosecution. However, cautioning against rapid introduction of SVA testimony is the fact that there is as yet no a standardized certification in SVA in the U.S.. Given the difficulty in learning how to use the technique, such rigorous certification is an essential prerequisite. Perhaps child protective service agencies and law enforcement agencies could begin to bring expert German counterparts to the U.S. to develop a consistent and uniform training program. The fact that lay evaluators often rate true statements as false speaks to the need for defense attorneys and prosecutors to both press for the adoption of this technique.

Dissertations generally conclude with a call for more research. Not here! That is to say, more research *is* called for, but categorically *not* of the sort of increasingly prevalent CBCA study that involves either adult statements or child statements of weak ecological validity. Although many respected authors have made passing references to SVA and called for more research, a careful reading of the international literature suggests that the few experiments which use real statements of abuse have demonstrated the technique to be exceptionally effective. Those who make passing reference to SVA often state that empirical studies have found conflicting results as to the accuracy of SVA or the reliability of expert application of the criteria for child sexual abuse statements. However, if one reads the literature carefully, it becomes apparent that those studies that cast doubt on the utility of the technique involve designs that limit the generalizability of their results to sexual abuse statements. This point cannot be overemphasized!

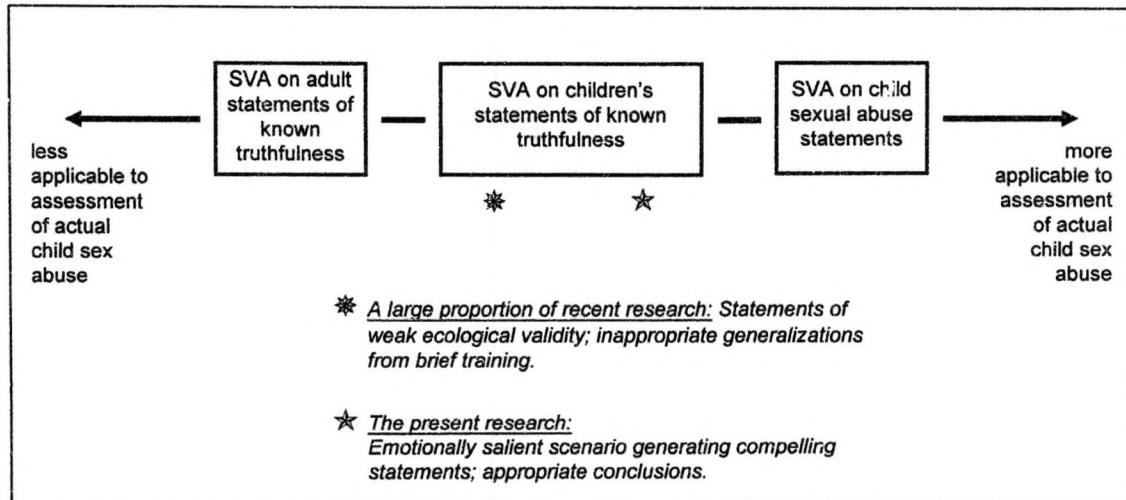
There are an increasing number of methodologically challenged designs in North American CBCA research. The term "training" has been used without qualification even though the most extensive training used in North American studies is but a fraction of what is done in Germany. Furthermore, briefly trained lay evaluators have often been asked to evaluate statements of adults, and the findings have been used by other authors to question the value of CBCA for evaluating children's statements. Even research that involves children's

statements is of limited value if the children have not been given emotionally compelling reason to lie (or, in truthful conditions, tell the truth).

Research may be visualized as falling on a continuum of applicability to child sexual abuse statement assessment. This continuum would be divided into three sections, as shown in Figure 9. On the left, with the weakest applicability to child sexual abuse assessment, are those studies that involve adult statements. In the middle section are experiments using non-sexual abuse children's statements of known truthfulness, and on the right are studies that use actual sexual abuse statements. While this type of field research is most strongly applicable to sexual abuse investigations, it always faces the challenge that the truthfulness of actual abuse statements cannot be determined.

Within the middle section of this continuum there is a great range of applicability (or generalizability) to actual sexual abuse investigations. Many laboratory studies are only weakly applicable—limited by their use of statements of poor quality or weak emotional validity—and make inappropriate generalizations from brief training programs. It would be unfortunate to draw conclusions about SVA's ability to determine the truthfulness of children's statements from such research. It is nevertheless possible to conduct sound laboratory experiments. One area that is especially appropriate for laboratory research is the impact of developmental factors and individual differences on CBCA.

Figure 9: Conceptual Continuum of SVA Research



We must remember Undeutsch's concerns about laboratory research and arrest the drift in experimental focus that has occurred. The best directions for future efforts in SVA involve keeping to the strong side of this applicability continuum. This would include detailing the current use of the technique in Germany, developing standardized training programs in North America, and further investigating the utility of SVA (including the SVA interview) when used by front-line professionals. New laboratory research on CBCA must focus on actual abuse statements and on laboratory generated statements of high saliency. Researchers must avoid easy experiments with statements of weak ecological validity, and reviewers must be wary of incautious generalizations. Statement Validity Assessment offers a powerful set of investigative and assessment tools that are greatly needed.

APPENDICES

APPENDIX A: SAMPLE STATEMENT

The reader is invited to assess the truthfulness of this statement. The last page of Appendix D reveals the answer.

- S: Someone took it. That, some guy. He, he was helping me and guess what he, that guy gave me?
- I: What?
- S: 50 cents. And he stole, he stole the book.
- I: He did? Why don't you sit down and tell me about this, <Subject>? Okay?
- S: He was, he was just helping me, we were doing cards and stuff. And I, and I had to weigh and he was seeing how fast my heart was beating and if I went very fast, I got 50 cents.
- I: Wow! That's great!
- S: And then, and then he was going to go and then he just, and that guy came right in and he, and he said "Have you guys seen my book, cause I was looking for it?" But, I think they were doing an experiment. I neat one. I'm gonna get that guy. That guy who stole the book, I'm gonna take my <unclear> and hold them, if I put it on his arm and I buckled it on and he lied on something, he gets a permanent shock.
- I: Oh. That would hurt, wouldn't it? Okay. Now I want you to tell me more about this book. You said somebody stole the book, can you tell me more about that?
- S: When we first got in, he was reading it, but when he came back in...gone.
- I: It was gone? What happened to it?
- S: I...I...I think that guy stole it.
- I: Which one?
- S: The guy who was helping me.
- I: Oh. Do you remember what his name was?
- S: What was it mom?
- I: Do you remember? It's okay if you don't remember.
- S: <shakes head no>
- I: Maybe you'll think of it. If you think of it you let me know, okay? So, okay, you said that, that this fellow that was helping you, he stole the book. Can you tell me more about that?
- S: When that, when that kid came back, he said, "I'm gonna call the cops." And he really did. And me and my mom thought we were, we thought we were supposed to, we didn't know we were supposed to go to the to the police station.
- I: It was a surprise, huh? Well, it's a serious thing when somebody steals a book. Can you tell me, <Subject>, everything that happened from the time that you arrived until you came over here? Can you remember and tell me everything that happened?

- S: First, when I just came in he was reading it, and the lady said that I'm gonna have a 'spearmint. So, they went, first, I went to stand on the scale and then he was going get my, he was doing this <shows action of taking the pulse> to see how fast my heart was beating. Then, and then he took me to the where the card games were and then, then, once I got to do it slow. But the 2nd time I had to do it super fast, and I did and I got these money. And then he went, he went, and that, and then he just came in and book was gone.
- I: It was gone, huh? Did you see what happened to the book?
- S: I saw, I saw him holding the book and and and just walked away.
- I: Then what happened?
- S: That's the whole thing I saw remember.
- I: Anything happen after you played the card game and you got the 50 cents?
- S: Then he just left.
- I: He left? Who left?
- S: That guy who was helping me.
- I: Okay. And then what happened?
- S: I can't remember that.
- I: Okay, just think.
- S: He didn't. I saw him taking the book home.
- I: Okay.
- S: He was walking down and then I <unclear>. That's the only thing I can remember.
- I: Where you in the room alone?
- S: Uh uh. My mom and a lady and that guy was there.
- I: Okay. Was your mom and that lady and everybody in the room all at the same time?
- S: Um hum. But not that guy, not that kid who was, he had, he's having a big test tomorrow. So he should, you know, get that book.
- I: But, can you remember the name of the guy who was working with you?
- S: <shakes head no>
- I: Can you tell me what he looked like?
- S: He looked like, he had a little beard, that <unclear> and he said to me. Now I remember one thing. After the card game he had me draw, but the only thing I know how to draw is pirate ships.
- I: That sounds like fun.
- S: That's what I want to be when I grow up... a pirate.
- I: Okay. So he had you draw? What happened after you drew?
- S: Then, then he talked to me and then then he went.
- I: Do you remember what he said to you when he talked to you?
- S: What my favorite show was, and what my favorite food was, and the rest.

- I: Can you remember anything else? You said somebody come back in the room.
- S: That same kid he was he was in there when I just came in, he was sitting down and reading. But all the books he found was his writing book. That's all. I found my writing book, mom. I found it. I have, I have a writing book at my house.
- I: Okay. That's good. Okay, there's something I'm gonna have to ask you because I'm a little bit confused. You said that, you said that you...Did you see the book when you first went into the room?
- S: Um hmm.
- I: Do you remember what the book looked like?
- S: Yep...I..uh uh <shakes head no>.
- I: You just think about it for a second.
- S: It had words. I couldn't read it because I'm in first grade, but I can't <unclear>. <unclear> only thing I remember.
- I: Do you remember what color it was?
- S: I remember it was gray. It was gray. And yellow. And brown. And red. It was all different colors.
- I: Okay. And then you started talking about this guy who was working with you. You said he had a beard. Can you tell me anything else about him?
- S: Um, um, I can't remember.
- I: Okay.
- S: Do you have any candy bars, have you got any candy machines anywhere?
- I: Well, not right in here, but maybe after a little while we'll see if we can find something for you.
- S: Because this is, I'm trying to same these things so I can get a candy. But at the school the candy machine, it didn't work. It never took one of my quarters. And we were, and my mom was trying to get that candy bar out. <unclear>
- I: Oh, darn it. Okay. Now. You said the guy whose book was gone, he came back into the room. What happened then?
- S: He just said, "Has anybody seen my book?" We never seen it and I was looking for it. I told my mom to get up so I could look. Hay, mom, remember that guy? He must've took the book and hided it right under the mattress or something I can't remember.
- I: Did you see somebody take the book?
- S: Huh uh. <shakes head no>
- I: No? You didn't see anybody take the book?
- S: Huh uh, all I, I just seed his writing book, that's all I saw.
- I: Okay, but you didn't see any, any, see where the book went?
- S: No.
- I: Okay. Can you remember anything else?
- S: Huh uh. <shakes head no>
- I: <debriefing>

APPENDIX B: EXPERT TESTIMONY

- I: Would you please state your name and address for the record?
- E: It's Charles Robert Honts. H-O-N-T-S. I live at twenty-two sixty three Springbrook Court, Grand Forks, ND.
- I: What is your occupation?
- E: I am an Associate Professor of Psychology at the University of North Dakota.
- I: What is your educational background?
- E: I have a bachelors degree in Experimental Psychology from Virginia Polytechnic Institute and State University in Blacksburg, Virginia. I received a Masters of Science in Psychology from that same institution, in 1982. I hold a Ph.D. in Experimental Psychology from the University of Utah which I received in 1986.
- I: Are you a clinical psychologist?
- E: No, I am an experimental psychologist.
- I: Can you work in the field of psychology without being a clinical psychologist?
- E: Absolutely. Clinical psychology is only a small part of the field of psychology.
- I: Do you concentrate your work in any particular area?
- E: Yes, over the last few years, my work has been concentrated in the area of psychology and the law. I am particularly interested in how psychology can offer techniques that may be of use in the legal system.
- I: Do you have any specialized training in this area?
- E: In several areas. I have been trained as a polygraph examiner since 1976, have done a fair amount of research in that area and have been to numerous advance seminars and workshops in that area. I have also been trained in techniques for interviewing and assessing the credibility of children, techniques that came to us from Germany, and I have been trained by the people who originated those techniques in Germany.
- I: Have you published any writings regarding credibility assessment?
- E: Quite a few. I have over a hundred publications and papers on credibility assessment with children and adults.
- I: Have you done any consultations for public agencies regarding credibility assessment?
- E: Yes. I worked with quite a few government agencies, agencies of the federal government. I have worked with the U.S. Secret Service, the Air Force Office of Special Investigations, the Naval Investigative Service, the Central Intelligence Agency. I also do quite a bit of work with the Royal Canadian Mounted Police in Canada.
- I: Do have any professional licenses?
- E: Yes, I am licensed as a polygraph examiner in the State of North Dakota.

- I: Do you belong to any professional associations?
- E: Quite a few. I am member, a charter member, of the American Psychological Society, which is the largest organization of experimental psychologists. I am a member of the Society for Psychophysiological Research, I am a member of American Psychology and Law Society, and I am a member of the Rocky Mountain Psychological Association.
- I: You said earlier that you were familiar with the credibility assessment of children's statements. Will you please explain to me what that is?
- E: Well, yes. There are a set of techniques that are relatively new in the English speaking world but which have been around in Germany and in Sweden for about thirty years. I became familiar with these techniques in the middle nineteen eighties. They are a set of techniques that are designed to be used with the child witness, in an effort to develop information about what the child witness knows and then to assess whether the child's statement is credible or not.
- I: How exactly does this process work?
- E: Well, the technique, which is known as Statement Validity Assessment has three phases. In the first phase an interview is conducted. It's a very special kind of interview, it's a structured interview, it's designed to raise the child to the child's highest level of functioning. It's designed specifically to develop information, it is not therapy. It's a forensic investigative interview. We know from decades of psychological research that the most accurate information can be obtained from statements that are given from free recall. And, so the interview is oriented to generate as much memory from recall as possible. And so the interview is designed to produce as long as possible a free narrative statement by the child. The interview then moves through a structure to progressively more and more pointed questioning and the entire approach is one of hypothesis testing. That is to say, initially the investigator who is doing the interview will form a series of alternative hypotheses about the case. For example, in a child sex abuse you might consider several hypotheses. One is obviously that the child is telling the truth, another might be that the child is basically telling the truth but has substituted a different perpetrator because the motivational situation is such that the actual perpetrator is too threatening, and so a less threatening person is named as the perpetrator. The child may have given a completely false statement because some other powerful adult has influenced the child to make a false statement. Yet another alternative hypothesis might be that the child for his or her own motives has come up with a false statement perhaps to get revenge or get even, or perhaps of because of psychological problems. So, working from this approach of having multiple hypotheses, the interview is designed to collect data to decide which of those hypotheses is the most likely.
- I: In this interview process what sort of questions do you start with?
- E: Well, the interview begins with very open ended questions, that is you might say, to the child that "I understand that there have been some problems at your home, can you tell me about that?" And so it's not suggested but it opens the stage and should begin to generate a free narrative. And the idea is to always ask very open ended questions and to lead the child as little as possible.
- I: Do you ever use any sort of leading questions in your interview process?
- E: Usually not. If there are important issues that are unresolved the questioning becomes progressively more focused. The only time we would use very leading or suggestive questions would be at the very end of the interview. If the interviewer has decided that one of the likely alternatives is that the child's accusations have been suggested by an adult, then you might try leading questions at that point just to get a feel for how suggestible this particular child is. That is not completely diagnostic because the child could be telling the truth and be suggestible, or the child could be lying and resist suggestion. But, it is suggestive.

- I: You said there were three phases. The first was the interview phase, what is the second phase?
- E: Yes, the second part is the statement analysis process. These interviews are always videotaped. Once the interview is completed, the interview will be transcribed and then there's a credibility assessment procedure that's engaged in with the transcribed interview. That is called criteria based content analysis, or C-B-C-A for short. What is done in CBCA is that you look for certain characteristics of a statement that are indicative of truthful statements. This is based on a hypothesis that was derived by a German psychologist in the nineteen fifties. And if we can, I'll go to our some of the materials that I have to help ease my talk along. [Undeutsch Hypothesis overhead displayed.] The German psychologist who came up with this notion is called Undeutsch. He is one of the psychologists who has trained me in this technique. It is based on a relatively simple notion that is known as the Undeutsch hypothesis. And that hypothesis says that statements derived from memories of actual experience differ in content and quality from statements based on invention or fantasy. The idea is that when we recount memories that are true memories, they will qualitatively differ than the kind of statements that we make if we are fantasizing or we are deliberately lying about what we have to say. The approach in CBCA is to look specifically for those criteria that are indicative of a truthful statement.
- I: What are some of these criteria?
- E: There are a number of them that we look for. They fall into three general categories. There are three that are basic general characteristics of statements. [Begin display of criteria overheads.] The first is one is that statement has logical structure. That is, does the statement make sense? That does not mean that the statement goes A-B-C-D-E-F, it just means that as we look back on it, does it make sense? Is it logical? Could it have happened? That is a basic requirement. If the statement doesn't have logical structure you're going to be very skeptical of it. The second criterion is called unstructured production. And, this one is a little bit counter intuitive. But, we know that when we recall things from memory, memory is not like a videotape. When we remember things that have happened to us it's not like we are playing a movie. We remember important things first and then as we are going through the process of remembering we remember other things, there are details. The order of memory is rarely A-B-C-D-E-F. It's far more often likely to be B-D-E-F-A-G. But, at the end if you look at the statement in it's entirety, we can piece it back together. On the other hand if we're telling a story that is not the truth, one of the things that you have to do when you are telling untruth is that you have to keep it straight. The basic way of keeping it straight is to organize it so that you tell it A-B-C-D-E-F, that way you don't get confused and you tell the story straight ahead. And so the second thing we look for in terms of general characteristics is that the statement be somewhat unstructured, that there be some moving back and forth in some inconsistency in the chronology in that sense that the person jumps around a little bit.
- I: Go ahead.
- E: The third general characteristic is that we look for quantity of detail. Our memories tend to be fairly rich we remember things not only central specific items, but we remember peripheral details, we remember how we felt, and that people's memories tend to be quite rich, in that you tend to see a fair production in detail associated with true statements.
- I: What are the other things that you look for?
- E: Well, we move on from the general characteristics of the statement to more specific contents. These are things that don't occur in every statement but when they occur they can be very powerful indicators that it is a truthful statement. And let me reemphasize that all of these criteria are things that are reinforcing that the statement is truthful. So these are things we expect to see in true memories. These criteria begin with one called contextual embedding. Everything that happens happens in a certain context. It happens in a place and at a certain time. And one of the things

that comes out when we tell about things we have actually experienced is we embed them in a specific place and time. We expect to see descriptions of interactions, and this has a very specific definition for us in this analysis in that it requires at least three elements. So in a statement we expect the child to say "He did this—I did that—he did something—I did something" before this criteria is met. We want to see the action move back and forth. And again this is something that is characteristic of true memory but does not seem to be very characteristic of stories that people make up. False stories typically have a one person kind of perspective "He did this—he did that—he did the other," rather than seeing the action move back and forth as it actually does in a true incident. Similarly the sixth one is reproduction of speech. Human beings are verbal, there is almost always conversation that goes on, even in instances of abuse. And, one of the things that happens when you hear a child talk about an actual experience is that they reproduce speech. So we look in the statement to find places where there should be quotation marks. "He said this, I said that, he said something else."

The next group I think are fairly self explanatory. We look for anything in the statement that is unexpected. If we're telling a story that is not the truth we are not going to complicate with unexpected things. And yet, in the real world unexpected things happen all the time. Some incident happens that interrupts one of the episodes of abuse. Some accident occurs, something unexpected will be in the statement—it's not the kind of thing you would make up to put in a statement if you weren't telling the truth. These don't happen all the time, but when they happen they are very persuasive. Unusual details are a similar thing. Unusual events are not the kind of things you would make up but nevertheless are things that are plausible and that make sense to us. Number nine, superfluous details, refers to information that are the kinds of things that get stuck in our memory but they're not necessary for the action of the story. Children often remember what they were wearing or they remember lighting in the room or the weather on that particular day. Those kinds of things don't really matter for retelling the story of abuse and you're unlikely to have added them if you were making that story up. But they are the very kinds of things that get attached to our real memories. Ten, another one I think that is fairly obvious, is accurately recorded details that are misunderstood. Again, these don't occur with high frequency, but when they occur they are very persuasive that you have a true statement. In this case the child describes an event accurately and to us as adults with sexual knowledge the descriptions make good sense, but the child clearly does not understand what has happened. And yet they have given us an accurate description of something that they clearly don't understand. Again, that could clearly be a very powerful indicator although it does not seem to happen with that much frequency. Related external associations would refer to events that happen in the context that show that it is a context of a sexual nature. For example, the discussion of sexual events that are not related directly related to the abuse. Perhaps the use of pornography. Other things that might not be added to if you were making the statement up but that support the accuracy of the statement by developing it. Number twelve is an important one. Our research shows that this one is very diagnostic. Human beings are verbal but we're also very affective. We have an important emotional life, and a big part of our memories is how we felt when certain things were going on. Sexual abuse is often traumatic for the children, they may be afraid, angry, anxious—those emotions get attached to the memory, to the real memories. And, one of the things that frequently comes out when we are talking about real memories is how we felt. And so, we see that that's a strong indicator that it's a truthful statement. The research indicates that this is rarely attached to false statements. Similarly, another thing we do as human beings is that we try to figure out what is going on in the minds of other people. We are constantly making attributions about why other people do things. One of the things that we find in true statements is that there will be attributions of the accused's mental state. That is the child will make attributions about why the perpetrator did the things that the perpetrator did. Again, this is not something that we expect to see if the child making up the story or has been coached to give a story.

The final category of criteria are called motivation related contents, and they have to do with things that a lying person is unlikely to do because they call attention to the fact that you might be lying. And yet they are things that people often do when they are recounting true memories, because memories are fallible and it is not unexpected that you will forget certain details, and children are often very nervous that they will not be believed. These criteria involve things like spontaneous corrections of the story so the child says something and realizes that he or she has made a mistake, and they go back and spontaneously fix the mistake. If you're lying, you are very unlikely to do that because it calls attention to the fact that you have made mistakes. An example of admitting lack of memory or knowledge would be a child saying "You know, I really want to remember what happened then but, I just don't remember that." This would be a spontaneous admitting a lack of memory. Simply answering a question, a direct question, with "I don't know" is not sufficient to meet this criterion. This criterion requires a spontaneous admission of a lack of memory or knowledge. People who have been abused, children who have been abused, often are concerned that they won't be believed. And, they will say things like "I know this sounds really bizarre, no one's going to believe this really happened to me, but it really did." Again, a person who is not telling the truth is unlikely to say those kind of things, because it calls attention to the fact that the story might not be truthful. Criterion seventeen is self-deprecation. Victims often blame themselves. Or on the other hand, if you're not telling the truth, you are unlikely to blame yourself. You're unlikely to call attention to yourself in that way. So if we see elements in the statement where the child partially blames themselves for this having occurred, that's taken as an indication that the statement likely to be credible. Another thing that often happens is that the child will often pardon the accused and try to make excuses for why the accused did the things the accused did, particularly if it's a family member, because the child may feel very uncomfortable about the fact that these proceedings may result in a family member being put in jail or taken out of the home or whatever. But on the other hand we don't expect to see that in a child who's making up a story. So all of these things these eighteen criteria are things we expect to see in a statement when the statement is true. It's not required that all of them be present, in fact it would be extremely rare for all of them to be present, but you want to see a preponderance of these things being present in order to conclude that a statement is truthful.

- I: Is there a scoring system for determining whether the statement is true?
- E: Yes there is. There is a check list that you have at hand when you're going through the transcript. As you find the criteria carefully reading through the transcript, you make notes. We use a format very much like lawyers do in preparing transcripts where the lines are numbered and so we'll write down the line numbers where these criteria occur and then we assign scores on a three point scale. A criterion can be either absent, present, or strongly present. So, if it's absent, that means we saw no sign that it occurred anywhere in the statement. If it occurs we assign a one. If we found it, and if it is a very strong occurrence that fits the definition for the criteria perfectly, then we would assign a two. So for each criterion there are three possible scores, Zero, if it's absent, One if it's present, or Two if it's strongly present. Even if a criterion occurs many times, it can't get a score higher than two.
- I: How many interviews should be performed on the child?
- E: Well, ideally only one. The Germans believe, and I think this is based in good science, that the first person to actually interview has the best chance of getting an accurate statement. Their approach is that if you do the job right the first time, and you have the documentation in the form of a recording of the interview, one interview is all that's necessary. And they feel that multiple interviewing is dangerous, and there is fair amount of scientific research to support that contention.
- I: Will the assessment be effective if it's the second or third interview of the child?

- E: That's difficult to know. That is a concern. There is some recent research from the University of Tennessee at Chattanooga that suggests that even after a series of misleading interviews, of suggestive and leading and very poor quality interviews, that if you do a good job of interviewing a child, it's still possible to get back to the actual memory. And so, I don't think that the science is closed on that issue, but I think it is suggestive that this approach works even after a couple of bad interviews.
- I: Is there anything else about the three phases?
- E: Well, we haven't talked about the third phase at all. [Checklist overhead displayed.] The third phase involves going back and looking at the entire interview and context of the statement in perspective. And that involves assessing a number of things that may have impacted on the quality of the CBCA. It also looks at the child's motives. Initially there are psychological characteristics of the child that you want to take a look at and take into consideration. Does the child have cognitive or emotional limitations that would have made it difficult for him or her to have given you a statement that would be high quality? So if the child is relatively nonverbal or extremely young, you would take that into account when you evaluate the results of the CBCA. On the other hand, if the child appears to be capable—that is, verbal and able to give detailed descriptions of things—and you get a low score on the CBCA, that's evidence that the CBCA is predictive. Did the child use language and knowledge that are appropriate for a child of that age and if they didn't, where would that language have come from? In some homes it's normal to use adult or medical like terms for the genitalia and so on. In other homes it would be very unusual, and that's something that you would take into consideration. What is the child's affect like? Was the child appropriately serious about matters that the child should have been serious about, or was there inappropriate affect displayed during the interview? Was the child flat about things that should have been very distressing? Was the child happy about things that should have made the child anxious? Those are all things you'd like to take into consideration. Is there evidence that the child is suggestible? And, if so, are there possible sources of suggestion in the child's case history? Was there an interview at one point that was highly suggestive? Is there an important adult figure who might possibly have led the child to say things that weren't true? So again, these are sort of checks on what you see coming out of the CBCA. We also want to ask some questions about the interview that you're assessing and also about other interviews. The interview that you're assessing—was it adequate? Does it meet the standards for applying this technique, that is, do you have free narrative? Was the interview leading or suggestive? Were there other interviews that were leading or suggestive that might have resulted in what you have in front of you when you do this particular analysis. So, those are things that we look at and that we would take into account. We also want to look at the motivational context of the accusation. That is, what are the motives for this child reporting? When you are assessing the alternative hypothesis, does this child have a motive for not telling the truth? Are there powerful adults in this situation who might have influenced this child? Does the child have something to gain by making a false statement? What was the context of the original disclosure? Is this something that happens spontaneously? For example, did the child go to one of its parents and say, "Something's happened to me, I'm frightened, so and so scared me" and then the accusation came out in that context? Or did it come out in the context of a heavy handed, leading and suggestive interview? So, we would want to take that into account, and of course the influence of others which I have already mentioned. Are there others who might have suggested, coached, or pressured the child to make a statement? And then finally, there are some investigative questions that we would want to take into consideration at the end of this process. Is the story that the child has told realistic? Are there elements in the story, important elements, that are simply not possible, that may violate the laws of nature? Are there contradictions in the statement or statements? And by this I am not concerned about contradictions about peripheral elements. We know that people remember things like colors and counts very poorly. Are there contradictions though about critical elements, elements of behavior that are important for the statement, things that are unlikely to

have been misremembered? And finally, does the statement have characteristics that are typical of this kind of offense? We know from many years of psychological research that there are certain elements that are characteristic of child sexual abuse. Usually there are threats, promises, secrets are kept, there are certain elements about progression, and so on. For example, incest typically begins with touching, and then progresses to more and more serious acts before intercourse actually takes place. It is very unusual in an incest situation for the first sexual contact to be full sexual intercourse. And so those are all things that we want to take into account when we evaluate the CBCA. The CBCA is not the final answer but it's an important part of the picture, and basically here we're looking to find out whether there is some reason to question the CBCA, or if we have elements that support its conclusions.

I: Is there anything else which should be mentioned regarding credibility assessment?

E: I think we've described the process fairly well.

I: Is this a new field?

E: It's a new field in the English speaking world. It's not a new field in psychology. It's been around in Germany since the early nineteen fifties. Undeutsch testified in German courts for the first time in, I believe it was in 1954. And they were lucky I think, in the sense that German courts looked at it, and it went all the way to the supreme court in Germany. They not only looked at it, but they mandated it. And so it's been a requirement in the German court system since 1955. On the other hand, since it was admissible and acceptable in Germany, in fact, since it has been mandated since 1955, there has been relatively little research. They didn't write about it very much and they didn't do a lot of research. So the research in the area is a fairly a new phenomena and primarily an English language phenomena.

I: Can this theory be tested?

E: Yes it can. In fact I've been involved in doing some of that validation research. We have done some of that at the University of North Dakota. And there are roughly two approaches to how you would go about testing this theory.

I: What are those two approaches?

E: Well, the two approaches that you can take I think are typical of science and how science goes about validating new ideas and notions. We can do things in the laboratory, and scientists like to work in the laboratory because in the laboratory we have good solid control over the things that we do. The other place that you do validation research is in the field. Obviously we want to know how accurate this is with children who have been sexually abused. We can't sexually abuse children in the laboratory for obvious ethical reasons and so we have to do other kinds of things in the laboratory.

I: What kinds of tests are these?

E: Well, there are several studies in the literature that have taken the field approach, these are not studies that I have done but they are studies that I am familiar with. I know the researchers who did the work that involve actual cases of child sexual abuse. The difficulty in doing that kind of research is in developing the criterion that tells you who was abused and who was not. And so what these investigators looked for was to find cases where the alleged perpetrators actually confessed to having committed the abuse. That's not a perfect criterion because occasionally innocent people do confess to crimes. But it's a pretty good one. And so you can look at those cases with a fair amount of confidence that the children in those cases are actually telling the truth when they say that this perpetrator abused them. Far more difficult is coming up with a criterion for false statements. One, because false statements are less frequent, much less frequent. And then two, because demonstrating that is not easy. And so the approach that's been taken in the published studies is one of using

multiple criteria. And in those studies they basically have required that two of three things be present. One would be that the child recants the allegation of abuse. And of course that is problematic because children do sometimes make false recantations. And so along with that they wanted the alleged perpetrator either to have taken and passed the polygraph test, or for there to have been a lack of physical evidence when there should have been physical evidence. So if there was an allegation of full penetration there should be some physical evidence of that and if there was none that was considered an element of the three that would be used as a criteria. Based on that research CBCA was shown to be an extremely good discriminator of supported versus questionable statements in both of those published studies.

At the University of North Dakota we have taken a somewhat different approach. Rather than working in the field where we have this difficulty with the criterion, we've decided to use the laboratory approach. And although we can't look specifically at child abuse, we've come up with a paradigm we think is very similar in many regards and that involves the child in an act that the child believes is real and is emotionally involving and involves serious matters. In a study we just finished recently we had three conditions. The first condition is a condition where the child is telling the truth. In that condition the children are told they are being brought to the University to take some psychological tests. They come with the parent, the parent is actually working for us at this point, because we recruited the parent to be part of our research team. There they are met by two psychologists, they go into a laboratory and the parent goes off with one of the researchers while the second researcher starts giving the child some tests. When they first arrive in this room, there is an unexpected complication. There is student in the room studying, and the researcher asks the student to leave. There is some discussion about the student who would like to study in that room—can the student come back later—the researchers agree and the student leaves all of his or her possessions in the room. So the child is being tested. We do some very simple psychological tests like sorting cards and drawing pictures. And, at some point the parent and researcher number one leave the room. The child and researcher two are left alone in the room together, and at some point researcher two goes over and looks through the student's possessions, finds a book—and we have this scripted so that everybody gets exposed to the same statements and same acts—and says that it's a nice book and that he or she is going to take it. And then he or she asks the child not to tell anyone. Mom and the other researcher come back, researcher two leaves with the book, a few minutes later the student returns, a big dramatic scene ensues, the student says "My book is gone! Where did my book go? It cost me forty-five dollars I've got an exam tomorrow, this is awful!" The student looks at the child and says, "Do you know who took my book?" Under those circumstances, with the perpetrator gone, most children will make an accusation at that point. They'll say that researcher two took the book, whoever that may have been that day. And, at that point the student goes, "This is just awful, I'm going to call the police." We had two waves of data collection, in the first wave of data collection we took the children and their parents over to the campus police station and we actually interviewed them at the police station. One of my graduate students did the interviews, but the children were led to believe that the person was a police officer. And so we did a statement analysis style interview, a very open ended interview designed to elicit as much free narrative as possible. There were two other conditions in this study that were designed to model, that is mimic the two kinds of false accusations that are made in the real world. In one of those conditions the book just disappears. No one knows what happens to it, it's just gone. The student returns, gets very upset, and in that condition walks up and accuses the parent of having stolen the book. The parent is left alone in the room with the child and looks at the child and says, "This is awful, they've called the police, they think I've stolen the book, I'm in terrible trouble, when the police come, why don't you tell them that researcher two took the book?" That's it, nothing else is said and the children aren't coached, they're just given a strong suggestion that they should falsely accuse researcher two. About two thirds of the children do falsely accuse researcher two in that condition. And they were then given a statement analysis interview, and they do give statements

about researcher two stealing the book. We think that that's a model of when a child just fantasizes a statement. So they just sort of make it up out of whole cloth. They haven't actually witnessed anything but they make it up as best they can. In the third condition, which we believe is a model for a perpetrator substitution sort of a setting, it's exactly like the first condition except that researcher two doesn't steal the book, the parent steals the book, and says exactly the same things that researcher two does—about how it's a nice book, that he or she will take it, and asks the child not to tell anyone. The parent then takes the book and tells the child that if anyone asks to say that researcher two took the book, so that the parent won't get in trouble. So the children have witnessed everything that the children in the true situation have witnessed, except there is a different perpetrator. Again, the student arrives, there's a scene, the police are called, and the children are interviewed.

We then took the statements that resulted from these three conditions, and we had people who were trained to do CBCA evaluate them without any knowledge of who was telling the truth and who wasn't. We also took the videotapes and the transcripts and showed them to a large number of average people and just asked them, "Just listen to this and can you tell us who is telling us the truth and who isn't?" The results of that study indicated: one, that the average person is not much better than chance at discriminating between true and false statements that are given under these circumstances. CBCA on the other hand performed quite well, based on the scorings that I described earlier, it correctly classified all of the false statements. It did make a few errors, and the few errors that it made were with children who were telling the truth, and there were a few children in that condition who didn't want to talk about it and gave very brief very impoverished statements. And so it made a few errors, the scientists would call them false positive errors, that is it falsely mistook some of the truthful children for not telling the truth. But the error rate is relatively small, about fifteen percent. One of my hats is as a statistician, and when we applied multivariate statistical techniques, we were able to get the accuracy rate up to a hundred percent. So at least in this laboratory setting, and under those control conditions, CBCA looks like a very powerful discriminator of truth telling and deception.

I: Has this been subject to peer review?

E: Yes it has.

I: And what was the result of that?

E: There are several published studies in the peer reviewed literature, there are two publications in the *Journal of Law and Human Behavior*, which is a first line psychology journal of the American Psychological Association. There is a publication in the *Journal of Behavioral Assessment* which again is again a first line publication, peer reviewed. And there have been a number of papers given at the American Psychology and Law Society, and the Society for Research in Child Development meetings, one of which I gave recently. And those are also peer reviewed before they're accepted.

I: Who would be the relevant scientific community for this field?

E: I believe that the relevant scientific community here would probably be the group of psychologists who belong to the American Psychology and Law Society, because they are scientists who are trained to evaluate scientific research but they also have an appreciation for legal issues, and they're very familiar with research in this area, because most of it has been published in psychology and law oriented journals.

I: Has this theory been generally accepted by this group?

E: I don't know of any surveys that have been taken, but I know at the last American Psychology and Law meetings there was no negative comment when these papers were given.

I: How many times have you testified in court as a psychologist?

E: Sixteen I believe.

I: How many times have you testified specifically on the subject of credibility assessment?

E: Twice.

I: How much are you compensated for your time?

E: For work that I do in my office in Grand Forks, I get a hundred and twenty five dollars an hour. And then when I travel out of town it's twelve hundred and fifty dollars a day.

I: Thank you, I have no further questions.

APPENDIX C: EXPERIMENTAL MATERIALS

<p>Consent Form (all materials have been reduced to fit)</p>
--

The goal of this study is to experimentally evaluate how well people can judge the truthfulness of statements of children. Some subjects may receive training in various techniques which we are looking at to see if this can improve the ability of subjects to evaluate the statements of children.

The ultimate benefit of this is to see if we can help ordinary people do a better job if they are ever on a jury.

You may first receive training in an evaluation technique by watching an expert testify in court. This training makes reference to certain characteristics of true and false statements, and may include reference to child sex abuse allegations. Some subjects receive no special training.

Then, you will watch videotapes and read transcripts of children making testimonies about the theft of a book. You will judge whether the child is telling the truth or not. Some of the testimonies contain true allegations about who stole a book, some contain false allegations.

Participating in this study will assist in important and interesting research in psychology and law, and will give you the chance to learn more about this topic as well as about experimental psychology. If you are uncomfortable at any time, you may leave.

Subjects may also, if they wish, compete for a \$50 prize based on the accuracy of their evaluations of the testimonies about the theft of a book. Additionally, subjects who were in training conditions may compete for another \$50 prize based on the accuracy of their memory of the training, measured after the evaluations of the testimonies are complete.

(Signature above certifies I am 18 years of age or over and have read and understood the above.)

If you have any questions, you may contact the investigator, Marcus Choi Tye, at 777-4348 or leave a message for him at 777-3451.

CBCA Testimony Handout

Feel free to write on this sheet. Note: please return it after finishing your evaluations.

CHARACTERISTICS OF A TRUE STATEMENT

Criteria for Analyzing General Characteristics

- Logical structure
- Unstructured production
- Quantity of details

Criteria for Analyzing Specific Contents

- Contextual embedding
- Descriptions of interactions
- Reproduction of conversation
- Unexpected complications during the incident

Criteria for Analyzing Peculiarities of Content

- Unusual details
- Superfluous details
- Accurately reported details misunderstood
- Related external associations
- Accounts of subjective mental state
- Attribution of perpetrator's mental state

Criteria for Analyzing Content Related to Motivation

- Spontaneous corrections
- Admitting lack of memory
- Raising doubts about one's own testimony
- Self-deprecation
- Pardoning the perpetrator

Criteria for Analyzing Offense-Specific Elements of Content

- Details characteristic of the offense

Counter-CBCA Testimony Handout

Feel free to write on this sheet. Note: please return it after finishing your evaluations.

CHARACTERISTICS OF A FALSE STATEMENT

Criteria for Analyzing General Characteristics

- Logical structure
- Unstructured production
- Quantity of details

Criteria for Analyzing Specific Contents

- Contextual embedding
- Descriptions of interactions
- Reproduction of conversation
- Unexpected complications during the incident

Criteria for Analyzing Peculiarities of Content

- Unusual details
- Superfluous details
- Accurately reported details misunderstood
- Related external associations
- Accounts of subjective mental state
- Attribution of perpetrator's mental state

Criteria for Analyzing Content Related to Motivation

- Spontaneous corrections
- Admitting lack of memory
- Raising doubts about one's own testimony
- Self-deprecation
- Pardoning the perpetrator

Criteria for Analyzing Offense-Specific Elements of Content

- Details characteristic of the offense

Verbal Instructions at Start of Session (from testimony condition)

"This is an experiment on credibility assessment. You will be watching a training video to help improve your ability to evaluate children's testimonies. After about 50 minutes of this you'll have a comfort break and then will evaluate some children's statements on videotape and in written form. All the children were in an experiment. In the statements the children will at some point accuse a research assistant of having taking a book. Either this accusation is true, and they saw the researcher assistant steal the book, or this is false. Your job will be to figure out which. Any questions? The experiment will take a little less than three hours. If you know you can't stay, we'd like to ask that you leave now. Thanks for your time and help."

Questionnaire, All Testimony Conditions
(from videotape statements first block)

CREDIBILITY ASSESSMENT RESEARCH

GENERAL DIRECTIONS:

Please put all your answers on the NCS Answer Sheet.

Be sure to check that you are filling in the "bubbles" in the correct places.

After a few minutes to answer general questions, you'll see some videotapes and then read some transcripts. Please evaluate each of the children's statements separately from the others.

PLEASE DO NOT MAKE ANY MARKS ON THIS QUESTIONNAIRE OR ON THE TRANSCRIPT PACKET WHICH WILL BE RE-USED.

OPTIONAL FOR SUBJECTS WISHING TO COMPETE FOR UP TO \$100:

If you wish to be considered for the prize money, put your name on the form in the NAME area, and put your telephone number in the IDENTIFICATION NUMBER area (no spaces or dashes). Be sure to fill in the corresponding bubbles. Your name will not be used in data analysis other than to figure out if you've won the prizes for the most accurate responses (a random draw will break any ties).

ALL SUBJECTS SHOULD ANSWER THESE QUESTIONS:

There are special areas on the left of the answer sheet for the answers to these three questions:

- Fill in the circle for **Sex**
- Fill in the circle for **Grade** (13-16, 16 for 5th year seniors, 0 if you are not an undergraduate)
- Fill in the circle for **Birth Date**

The remaining questions are numbered from 1 and should be answered in the row on the answer sheet with the same number. The letter designates which circle you should darken for a particular response.

1. **Have you ever served on a jury?**
 A - Yes
 B - No

2. **How many children do you have?**
 A - 0
 B - 1
 C - 2
 D - 3
 E - more than 3

3. **Please think about how many hours you've had in direct contact with children between the ages of two and twelve. This would include baby sitting, taking care of your own children, etc. Only count time spent with your own brothers and sisters if you were at least fourteen when they were age 2-12.**
 A - None
 B - Less than 20 hours ever
 C - Between 20 and 100 hours
 D - Over 100 and less than 500 hours
 E - Over 500 hours

4. **Have you ever received training in assessing the credibility of children before?**
 A - Yes
 B - No

5. **"A child makes an accusation about a crime to a police officer." Which *one* of the following statements *best* describes your attitude to this situation?**
 A - Children will always tell the truth about serious matters
 B - Children will usually tell the truth about serious matters
 C - Children will sometimes tell the truth about serious matters
 D - Children cannot be trusted about serious matters
 E - It depends on the child
 F - It depends on the situation

STOP.

The experimenter will soon show you videotaped statements and then printed statements for you to evaluate quietly by yourself.

In each statement at some point a child will make an accusation about a research assistant stealing a book. Either this is true (it really happened) or this is false (the research assistant didn't take the book). **Remember to evaluate each statement separately, since you've been assigned statements randomly and there is no fixed ratio of true and false statements!**

FIRST VIDEOTAPE

6. In the videotape you've just watched, at some point a child said that a research assistant stole someone's textbook. Please evaluate the truthfulness of the child's accusation:

A - The child was being truthful, the researcher took the book
 B - The child was lying and knows the researcher didn't take the book
 C - The child believed the researcher took the book, but the researcher really didn't take the book

7. How confident are you in this evaluation?

A - Completely confident
 B - Mostly confident
 C - Somewhat confident
 D - Fifty-fifty
 E - Somewhat uncertain
 F - Mostly uncertain
 G - Completely uncertain

SECOND VIDEOTAPE

8. In the videotape you've just watched, at some point a child said that a research assistant stole someone's textbook. Please evaluate the truthfulness of the child's accusation:

A - The child was being truthful, the researcher took the book
 B - The child was lying and knows the researcher didn't take the book
 C - The child believed the researcher took the book, but the researcher really didn't take the book

9. How confident are you in this evaluation?

A - Completely confident
 B - Mostly confident
 C - Somewhat confident
 D - Fifty-fifty
 E - Somewhat uncertain
 F - Mostly uncertain
 G - Completely uncertain

THIRD VIDEOTAPE

10. In the videotape you've just watched, at some point a child said that a research assistant stole someone's textbook. Please evaluate the truthfulness of the child's accusation:

A - The child was being truthful, the researcher took the book
 B - The child was lying and knows the researcher didn't take the book
 C - The child believed the researcher took the book, but the researcher really didn't take the book

11. How confident are you in this evaluation?

A - Completely confident
 B - Mostly confident
 C - Somewhat confident
 D - Fifty-fifty
 E - Somewhat uncertain
 F - Mostly uncertain
 G - Completely uncertain

FOURTH VIDEOTAPE

12. In the videotape you've just watched, at some point a child said that a research assistant stole someone's textbook. Please evaluate the truthfulness of the child's accusation:

A - The child was being truthful, the researcher took the book
 B - The child was lying and knows the researcher didn't take the book
 C - The child believed the researcher took the book, but the researcher really didn't take the book

13. How confident are you in this evaluation?

A - Completely confident
 B - Mostly confident
 C - Somewhat confident
 D - Fifty-fifty
 E - Somewhat uncertain
 F - Mostly uncertain
 G - Completely uncertain

FIRST TRANSCRIPT

14. In the transcript you've just read, at some point a child said that a research assistant stole someone's textbook. Please evaluate the truthfulness of the child's accusation:

A - The child was being truthful, the researcher took the book
 B - The child was lying and knows the researcher didn't take the book
 C - The child believed the researcher took the book, but the researcher really didn't take the book

15. How confident are you in this evaluation?

A - Completely confident
 B - Mostly confident
 C - Somewhat confident
 D - Fifty-fifty
 E - Somewhat uncertain
 F - Mostly uncertain
 G - Completely uncertain

SECOND TRANSCRIPT

16. In the transcript you've just read, at some point a child said that a research assistant stole someone's textbook. Please evaluate the truthfulness of the child's accusation:

A - The child was being truthful, the researcher took the book
 B - The child was lying and knows the researcher didn't take the book
 C - The child believed the researcher took the book, but the researcher really didn't take the book

17. How confident are you in this evaluation?

A - Completely confident
 B - Mostly confident
 C - Somewhat confident
 D - Fifty-fifty
 E - Somewhat uncertain
 F - Mostly uncertain
 G - Completely uncertain

THIRD TRANSCRIPT

18. In the transcript you've just read, at some point a child said that a research assistant stole someone's textbook. Please evaluate the truthfulness of the child's accusation:

A - The child was being truthful, the researcher took the book
 B - The child was lying and knows the researcher didn't take the book
 C - The child believed the researcher took the book, but the researcher really didn't take the book

19. How confident are you in this evaluation?

A - Completely confident
 B - Mostly confident
 C - Somewhat confident
 D - Fifty-fifty
 E - Somewhat uncertain
 F - Mostly uncertain
 G - Completely uncertain

FOURTH TRANSCRIPT

20. In the transcript you've just read, at some point a child said that a research assistant stole someone's textbook. Please evaluate the truthfulness of the child's accusation:

A - The child was being truthful, the researcher took the book
 B - The child was lying and knows the researcher didn't take the book
 C - The child believed the researcher took the book, but the researcher really didn't take the book

21. How confident are you in this evaluation?

A - Completely confident
 B - Mostly confident
 C - Somewhat confident
 D - Fifty-fifty
 E - Somewhat uncertain
 F - Mostly uncertain
 G - Completely uncertain

STOP.

Please wait quietly for the experimenter to hand you the final questionnaire.

Final Questionnaire, CBCA and
Counter-CBCA Conditions

Please put your name here: _____

this will only be used to match this page with your Scantron

PLEASE ANSWER #22-26 ON THE NCS FORM BY FILLING IN THE CIRCLES...

22. **Which made your evaluations of the children's truthfulness easier?**
 A - it was easier to figure out from the printed statements
 B - it was easier to figure out from the videos
 C - they were about the same
23. **How did CBCA training affect your ability to evaluate the children's statements?**
 A - It helped tremendously
 B - It helped a lot
 C - It helped a little
 D - It neither helped nor hurt
 E - It hurt a little
 F - It hurt a lot
 G - It hurt tremendously
24. **How competent was Dr. Honts?**
 A - very competent
 B - somewhat competent
 C - a little competent
 D - neither competent nor incompetent
 E - a little incompetent
 F - somewhat incompetent
 G - very incompetent
25. **How convincing was Dr. Honts?**
 A - very convincing
 B - somewhat convincing
 C - a little convincing
 D - neither convincing nor unconvincing
 E - a little unconvincing
 F - somewhat unconvincing
 G - very unconvincing
26. **The presence in a statement of CBCA criteria suggests that...**
 A - The statement is most likely true
 B - The statement is most likely false

ANSWER THE FOLLOWING BY WRITING ON THIS PAGE IN THE SPACE BELOW...

Have you had any prior exposure to this research, or has anyone talked about it with you?

Please circle one: **Yes / No**

If yes, we'd appreciate you letting us know what you've already heard before the experiment began:

Please turn this form over and answer the questions on the other side. *Thank you.*

Below, write down as many of the criteria as you can remember:

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

Please use the space below to write down any comments you may have about the research. We're especially interested in any other techniques you may have used to figure out which children were telling the truth. Thanks again for your time.

Final Questionnaire, No Testimony Condition
--

Please put your name here: _____

this will only be used to match this page with your Scantron

ANSWER THE FOLLOWING QUESTION ON THE NCS FORM...

22. Which made your evaluations of the children's truthfulness easier?

- A - it was easier to figure out from the printed statements
- B - it was easier to figure out from the videos
- C - they were about the same

ANSWER THE FOLLOWING BY WRITING ON THIS PAGE IN THE SPACE BELOW...

Have you had any prior exposure to this research, or has anyone talked about it with you?

Please circle one: **Yes / No**

If yes, we'd appreciate you letting us know what you've already heard before the experiment began:

Please use the space below to write down any comments you may have about the research. We're especially interested in any other techniques you may have used to figure out which children were telling the truth. *Thanks again for your time.*

APPENDIX D: RESAMPLING STATISTICS

Resampling offers a variety of statistical tools which are complementary to the traditional tests employed by researchers. This brief introduction to resampling makes reference to the tests conducted in the present study. Following this introduction are sample programs executed to find bootstrap probabilities reported on in the Results chapter.

One way of testing whether two observed frequencies differ significantly would be to perform a probability calculation, like the nonparametric BINOMIAL procedure in SPSS (comparing to a population with a mean frequency of .5). Resampling is analogous, but instead of calculating a probability using a formula, one finds it “experimentally” by conducting a Monte Carlo simulation. Resampling involves combining actual experimental groups into a single vector—saying that all observations come from the same population is a restatement of the null hypothesis that treatments have no effect on the group means—and then creating new experimental “observations” from this vector. The statistic of interest is calculated for these new observations (such as a difference between means), and this is stored. The process is repeated several hundred times, building up a sampling distribution of the statistic of interest. After this, one examines where the actual experimental value of the statistic falls

in the sampling distribution, and this indicates the (bootstrap) probability of having found results that or more extreme under assumption that the null hypothesis is true (Simon, 1992).

Using resampling to ask whether an observed frequency differs from chance requires generating a vector of random numbers distributed as one would expect under the null (e.g., 50% 0's and 50% 1's). Resampling makes it easy to directly compare two frequencies which result from different numbers of observations, without having to use intermediate statistics such as Fisher's Z' (Simon, 1992). It also makes it possible to find the power of a particular experimental design.

Generally, 1,000 trials is satisfactory. More trials results in a smoother distribution, but the incremental gain with each additional trial (in terms of improving closeness of fit to a mathematically exact probability) becomes ever smaller (Simon, 1992). The resampled probability thus derived will be very close to an exact probability (such as a Fisher's Exact Test), if a formula exists to calculate one. When no formula exists, resampling offers the only way of calculating statistical significance.

Thus, an advantage of resampling for the present study is that it allows direct comparisons of frequencies with different n (e.g., comparing 65% of 1000 lay evaluations correct with 89% of 28 expert CBCA evaluations). Although this is a simple question, conventional statistics do not readily apply to such a

situation. One could use an ANOVA, but the data are nonparametric (violating an assumption of ANOVA), and furthermore, ANOVA may be biased with vastly unequal n . One might also use Fisher's Z' to compare two correlations based on different group sizes, however, Fisher's Z' works with intermediate values (comparing two correlations) and is neither conceptually nor numerically simple to apply. Also, correlations again make assumptions about the underlying distribution of the population, which may not be true for the present data.

Resampling does not rely on any *theoretical* (underlying) distribution. By definition, it creates the null-hypothesis distribution given the data (which, under the null, come from the same population). By definition it is free from concerns about departing from an assumed population distribution, unlike probability tests (Simon, 1992). Furthermore, resampling allows questions to be answered that would be difficult to address with conventional statistics, or would require the misapplication of conventional statistics. In conclusion, as Simon (1992) notes, resampling statistics should be viewed as complementary to conventional techniques.

For the present research, resampling was used to conduct a power analysis (reported in the Method chapter). It was also used in comparing differential evaluation performance (reported in the Results chapter). The rest of this appendix contains resampling programs written by the author in the Resampling Stats language, which is a simple programming language designed

to construct resampling-style Monte Carlo analyses. The programs have been annotated to explain the function of the code and the program. Table 14 in the Results chapter compares differential evaluation performance to chance.

Following the programs is a list of the actual ratios (observed values) that were resampled to determine these bootstrap probabilities.

```

' Program to calculate a sampling distribution for the number correct under
' under chance conditions. Then finds the mean and standard deviation of
' this sampling distribution, and calculates number correct for  $z=1.65$ 
' (one-tailed  $p<.05$ ) (training condition, 1260 evaluations). [DISSPWR1.STA]
' equal to .1 correlation, 1% of variance, performance of 55% correct in
' experimental hypothesis population (vs. 50% correct under chance in null
' hypothesis population). 1260 evals = medium effect with 15 Ss per group
' and 55% of 1260 is 693.
' (4 evals per medium * 7 groups * 3 training conditions * 15 Ss/run = 1260)
' (8 evals per training cond x 7 groups x 15 Ss/run = 840 for main eff. train)
' (and 50% of main eff. training for interaction, 420 and 55% of 420)

```

MAXSIZE DEFAULT 10000

REPEAT 10000

```

GENERATE 1260 0,1 A ' generate 1260 random integers 0 or 1
COUNT A = 1 B      ' count how many 1's (correct), put into B
SCORE B Z           ' Z will be a sampling dist. of # correct

```

END

```

MEAN Z M           ' find mean of sampling distribution
SIDEV Z S          ' find standard deviation of sampling distribution

```

```

MULTIPLY S 1.65 A   ' one-tailed  $p<.05$ ,  $z=1.65$ 
ADD A M C

```

```

PRINT M S
PRINT C            ' bootstrapped # correct by chance at .05 1-tail limit

```

```

' any more than this correct by chance would be significant at  $p<.05$ 
' level (one-tailed). This came to 659 under when the program was run
' (mean 630, SD 17.771, 1.65 SD is 1-tailed  $z$  of .05)

```



```
' Program to calculate power of finding significance for training condition
' if effect accounts for 1% of variance in criterion (.1 correlation, or
' a performance of 55% correct). [DISSPWR2.STA]
```

```
' This uses sampling distribution of program DISSPWR1, which for 1260
' evaluations yielded 659 as the .05 cutoff (if more than 659 of 1260 are
' correct,  $p < .05$  for getting results that or more extreme if the null
' hypothesis is true).
```

```
MAXSIZE DEFAULT 1500
```

```
GENERATE 55 1 A
```

```
GENERATE 45 0 B
```

```
CONCAT A B C      ' Experimental Hypothesis population, 55% correct
```

```
REPEAT 1000
```

```
    SAMPLE 1260 C D
```

```
    COUNT D =1 E
```

```
    SCORE E Z
```

```
' generate a sample
```

```
' count how many correct
```

```
' create a sampling distribution of number correct
```

```
END
```

```
COUNT Z >= 659 X      ' given the experimental hypothesis (population
' is 55% correct), count number of times a result
' was found that was equal or more extreme than
' the number correct that would be found at the .05
' level given the null hypothesis
```

```
DIVIDE X 1000 Y      ' express as a proportion
```

```
PRINT Y              ' this is the bootstrap power
```

'Program which compares an outcome to chance.

'This program calculates the probability of getting \geq a particular number
'of observations correct under the assumption that the null hypothesis
'is true (i.e. that the population from which the group came is
'one in which performance is at 50% correct, i.e. chance level).

MAXSIZE DEFAULT 5000 Z 11000

GENERATE 5000 0,1 A 'make a large vector of random observations
SHUFFLE A A 'shuffle for good measure!

REPEAT 10000

'sample to however many actual observations there were (here, 2664)

SAMPLE 2664 A K 'sample with replacement to make an experimental "run"
COUNT K = 1 KK 'count number correct that was observed that run
SCORE KK Z 'preserve this across iterations, this will accumulate
 'to form a sampling distribution under the null

END

'set this equal to actual number correct

COUNT Z \geq 1731 ZZ 'count number of times a result equal to or greater than
 'that which was actually observed was obtained in the
 '10,000 Monte Carlo trials (here, 65% correct)
 'this would be \leq 1731 if we were looking for
 'performance being worse than chance

DIVIDE ZZ 10000 P 'express as a proportion (probability)

PRINT P 'actual probability given the null (chance performance)
 'of having obtained results equal to or more extreme
 'than were actually observed (one-tailed)
 '(here, one tailed: considering better than chance
 ' OR worse than chance only)

'Resampling program to compare any two frequencies/proportions correct.

'Program to calculate probability of finding a proportion correct
'that or more extreme than the one actually observed, given the
'null hypothesis that the observed group comes from the same
'population as a comparison group.

'Comparison group can chance (equal numbers right and wrong),
'or the numbers right and wrong in another experimental condition.

'BENEFIT OF RESAMPLING: Number of evaluations in the two groups may
'be dissimilar and comparison is still valid.

'This example will compare observation of 65% of 2664 evaluations correct,
'with a second observation of 90% of 28 evaluations correct.

MAXSIZE DEFAULT 3000 C 5500 Z 12000

'A1 (A2) -- observed number correct (incorrect) for group A

SET 1732 1 A1 'generate observed number correct (2664*.65)
SET 932 0 A2 'generate observed number incorrect (2664-1732)
CONCAT A1 A2 A 'pull together: A has actual data from first group
SHUFFLE A A

'B1 (B2) -- observed number correct (incorrect) for group B

SET 25 1 B1 'observed number correct (28*.9)
SET 3 0 B2 'observed number incorrect (28-25)
CONCAT B1 B2 B 'actual data from second group
SHUFFLE B B

CONCAT A B C 'under null, both groups come from same population
SHUFFLE C C 'shuffle again for good measure!

(continued on next page)

REPEAT 10000 'we will pretend we had run the experiment 10,000 times

'sample & divide by same number as total observations in group A

SAMPLE 2664 C K 'a single Monte Carlo run for first group
 COUNT K = 1 K1 'count how many were observed to be correct
 DIVIDE K1 2664 KK 'express # right as a proportion
 '(so we can compare with 2nd group)

'ditto for total observations in group B

SAMPLE 28 C L 'a single Monte Carlo run for the second group
 COUNT L = 1 L1 'count how many were observed to be correct
 DIVIDE L1 28 LL 'express # right as a proportion
 '(so we can compare with 1st group)

SUMABSDEV KK LL J 'express difference between 1st and 2nd group's proportions
 'as an absolute deviation (absolute value = 2 tailed)
 '(2-tailed because 1st could be > 2nd or vice-versa)

SCORE J Z 'Z will accumulate this difference, building our sampling
 'distribution of differences for our 10,000 "experiments"

END

'use difference in observed proportions:

' ([observed correct in A]/[total in A]) - (ObsCorrB/TotalB)

COUNT Z >= .25 ZZ 'how often in the sampling distribution we observed a
 'difference in proportion greater than or equal to the
 'actual observed difference
 '(.65-.9=.25, also 1731.6/2664 - 25.2/28 = .25)

DIVIDE ZZ 10000 P 'proportion (probability or p value) of times we observed it

PRINT P 'p value of obtaining a difference between the two groups
 'equal to or more extreme than what we actually observed
 'given the null is true and the two groups come from the
 'same population

**Actual Ratios Compared (Observed Values) for Resampled Bootstrap
Probabilities of Evaluations-Compared-to-Chance Reported on in Results**

EVALUATION	STATEMENTS		
	% correct exact ratio resampled (observed frequencies)		
	All	True	False
Lay Evaluators, No Training	<u>65%</u> 645/1000	<u>63%</u> 278/440	<u>66%</u> 367/560
Lay Evaluators, CBCA Training	<u>63%</u> 487/768	<u>64%</u> 211/328	<u>63%</u> 276/440
Lay Evaluators, Counter Training	<u>59%</u> 529/896	<u>47%</u> 173/372	<u>68%</u> 356/524
Expert CBCA, Discriminant Analysis	<u>89%</u> 25/28	<u>75%</u> 9/12	<u>100%</u> 16/16
Expert CBCA, Jackknifed Disc. Analysis	<u>71%</u> 20/28	<u>58%</u> 7/12	<u>81%</u> 13/16
Expert CBCA, Logistic Regression	<u>86%</u> 24/28	<u>83%</u> 10/12	<u>88%</u> 14/16
Expert CBCA, Jackknifed Log. Regression	<u>79%</u> 22/28	<u>67%</u> 8/12	<u>88%</u> 14/16

The child's statement in Appendix A is from a Fantasy Accusation condition, and is false.

REFERENCES

- Anson, D. A., Golding, S. L., & Gully, K. J. (1993). Child sexual abuse allegations: Reliability of criteria-based content analysis. *Law and Human Behavior*, 17, 331-341.
- Bekerian, D. A., & Dennett, J. L. (1995). Assessing the truth in children's statements. In T. Ney (Ed.), *True and false allegations of child sexual abuse: Assessment and case management* (pp. 163-175). New York: Brunner/Mazel.
- Bixler, R. H. (1992). Do we/Should we behave like animals? In W. O'Donohue & J. H. Geer (Eds.), *The sexual abuse of children: Theory and research* (pp. 81-107). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Box, H. O. (1984). *Primate behaviour and social ecology*. London: Chapman and Hall.
- Boyчук, T. D. (1991). Criteria-based content analysis of children's statements about sexual abuse: A field-based validation study (Doctoral dissertation, Arizona State University, 1991). *Dissertation Abstracts International*, 52, 2272A.
- Brainerd, C. J., & Reyna, V. F. (1995). Autosuggestibility in memory development. *Cognitive Psychology*, 28, 65-101.
- Brodie, L. A. (1993). *Making judgments regarding child sexual abuse: The impact of professional background, experience and knowledge base* (Doctoral dissertation, Biola University, 1993). *Dissertation Abstracts International*, 54, 3334B.
- Bulkley, J. A. (1992). Major legal issues in child sexual abuse cases. In W. O'Donohue & J. H. Geer (Eds.), *The sexual abuse of children: Theory and research* (pp. 139-167). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Ceci, S. J., & Bruck, M. (1993). Suggestibility of the child witness: A historical review and synthesis. *Psychological Bulletin*, 113, 403-439.
- Ceci, S. J., & Bruck, M. (1995). *Jeopardy in the courtroom: A scientific analysis of children's testimony*. Washington, DC: American Psychological Association.

- Ceci, S. J., Ross, D. F., & Toglia, M. P. (Eds.). (1987). *Children's eyewitness memory*. New York: Springer-Verlag.
- Corder, B. F., & Whiteside, R. (1988). A survey of juror's perception of issues related to child sexual abuse. *American Journal of Forensic Psychology*, 6, 37-43.
- Davenport, W. H. (1992). Adult-child sexual relations in cross-cultural perspective. In W. O'Donohue & J. H. Geer (Eds.), *The sexual abuse of children: Theory and research* (pp. 73-80). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Devitt, M. K., Honts, C. R., Gillund, B. E., Amato, S. L., Peters, D. P., & Norton, M. (1994, March). *A study of the willingness of children to make false accusations about a serious matter in a realistic setting*. Paper presented at the meetings of the American Psychology - Law Society, Santa Fe, New Mexico.
- deYoung, M. (1992). Credibility assessment during the sexual abuse evaluation. In W. O'Donohue & J. H. Geer (Eds.), *The sexual abuse of children: Clinical issues* (pp. 256-282). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Doris, J. (Ed.). (1992). *The suggestibility of children's recollections: Implications for eyewitness testimony*. Washington, DC: American Psychological Association.
- Doris, J. (1994). Commentary on criteria-based content analysis. *Journal of Applied Developmental Psychology*, 15, 281-285.
- Faller, K. C. (1993). *Child sexual abuse: intervention and treatment issues*. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Administration on Children, Youth and Families, National Center on Child Abuse and Neglect.
- Goodman, G. S., & Clarke-Stewart, A. (1992). 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, DC: American Psychological Association.
- Gordon, B. N., Ornstein, P. A., Nida, R. E., Follmer, A., Creshaw, C., & Albert, G. (1993). Does the use of dolls facilitate children's memory of visits to the doctor? *Applied Cognitive Psychology*, 7, 459-474.

- Hembrooke, H., & Ceci, S. J. (1995). Traumatic memories: Do we need to invoke special mechanisms? *Consciousness and Cognition an International Journal*, 4, 75-82.
- Honts, C. R. (1994). Assessing children's credibility: Scientific and legal issues in 1994. *North Dakota Law Review*, 70, 879-903.
- Honts, C. R., Devitt, M. K., Tye, M. J. C., Peters, D. P., & Vondergeest, L. (1995, March). *Credibility assessment with children*. Paper presented at the meeting of the Society for Research in Child Development, Indianapolis, IN.
- Honts, C. R., & Perry, M. V. (1992). Polygraph admissibility: Changes and challenges. *Law and Human Behavior*, 16, 357-379.
- Horowitz, S. W. (1991). Empirical support for statement validity analysis. *Behavioral Assessment*, 13, 293-313.
- Joffe, R. D. (1992/1994). Criteria-based content analysis: An experimental investigation with children (Doctoral dissertation, University of British Columbia, 1992). *Dissertation Abstracts International*, 54, 4439B [1994]. (Canadian Theses ISBN 0-315-80809-8.)
- Kinsey, A. C., Pomeroy, W. P., & Martin, C. E. (1948). *Sexual behavior in the human male*. Philadelphia: W. B. Saunders.
- Kinsey, A. C., Pomeroy, W. P., Martin, C. E., & Gebhard, P. H. (1953). *Sexual behavior in the human female*. Philadelphia: W. B. Saunders.
- Klitzing, K. V. (1990). Credibility examination of children and adolescents on the question of sexual abuse. *European Journal of Child and Adolescent Psychiatry, Acta Paedopsychiatrica*, 53, 181-190.
- Köhnken, G., Schimossek, E., Aschermann, E., & Höfer, E. (1995). The cognitive interview and the assessment of the credibility of adult's statements. *Journal of Applied Psychology*, 80, 671-684.
- Köhnken, G., & Steller, M. (1988). The evaluation of the credibility of child witness statements in the German procedural system. In J. Shapland & J. Drinkwater (Eds.), *The child witness - Do the courts abuse children?* (pp. 37-45). Leicester, England: The British Psychological Society for the Division of Criminological and Legal Psychology.

- Köhnken, G., Thurer, C., & Zoberbier, D. (1994). The cognitive interview: Are the interviewers' memories enhanced, too? *Applied Cognitive Psychology*, 8, 13-24.
- Lamb, M. E. (1994). The investigation of child sexual abuse: An interdisciplinary consensus statement. *Expert Evidence*, 2, 151-156.
- Lamb, M. E., Sternberg, K. J., & Esplin, P. W. (1994). Factors influencing the reliability and validity of statements made by young victims of sexual maltreatment. *Journal of Applied Developmental Psychology*, 15, 255-280.
- Landry, K. L., & Brigham, J. C. (1992). The effect of training in criteria-based content analysis on the ability to detect deception in adults. *Law and Human Behavior*, 16, 663-676.
- Memon, A., Cronin, O., Eaves, R., Bull, R. (1993). The cognitive interview and child witnesses. *Issues in criminological and legal psychology*, 20, 3-9.
- Newton, D. E. (1992). Homosexuality and child sexual abuse. In W. O'Donohue & J. H. Geer (Eds.), *The sexual abuse of children: Theory and research* (pp. 329-358). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Ornstein, P. A., Follmer, A., & Gordon, B. N. (1995, March). *The influence of dolls and props on young children's recall of pediatric examinations*. Paper presented at the meeting of the Society for Research in Child Development, Indianapolis, IN.
- Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change*. New York: Springer Verlag.
- Poole, D. A. (1995a, March). *Recent challenges to three commonly held assumptions about children's eyewitness testimony*. Paper presented at the meeting of the Society for Research in Child Development, Indianapolis, IN.
- Poole, D. A. (1995b, March). *Toward a model of developmental changes in eyewitness suggestibility*. Paper presented at the meeting of the Society for Research in Child Development, Indianapolis, IN.
- Quinn, K. M. (1988). The credibility of children's allegations of sexual abuse. *Behavioral Sciences and the Law*, 6, 181-199.

- Raskin, D. C., & Esplin, P. W. (1991). Statement validity assessment: Interview procedures and content analysis of children's statements of sexual abuse. *Behavioral Assessment*, 13, pp. 265-291.
- Raskin, D. C., & Esplin, P. W. (1992a). Assessment of children's statements of sexual abuse. In J. Doris (Ed.), *The suggestibility of children's recollections: Implications for eyewitness testimony* (pp. 153-164). Washington, DC: American Psychological Association.
- Raskin, D. C., & Esplin, P. W. (1992b). Commentary: Response to Wells, Loftus, and McGough. In J. Doris (Ed.), *The suggestibility of children's recollections: Implications for eyewitness testimony* (pp. 172-176). Washington, DC: American Psychological Association.
- Raskin, D. C., Esplin, P. W., & Craig, R. A. (1995, January). *Statement Validity Assessment*. (Available from the Department of Psychology and Division of Continuing Education, University of Utah, Salt Lake City, UT 84112).
- Raskin, D. C., & Honts, C. R. (1994). A bootstrap reanalysis of the results of Boychuk, 1991. In T. Skjæveland (Ed.), *Vitnepssykologi - 94: Psychological methods in the investigation and court treatment of sexual abuse*. Tromsø, Norway: University of Tromsø.
- 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 children's testimony* (pp. 184-207). New York: Springer-Verlag.
- Ruby, C. L. (1995). The utility of the criteria-based content analysis in distinguishing between truthful and fabricated statements made by white and black speakers (Doctoral dissertation, Florida State University, 1990). *Dissertation Abstracts International*, 56, 2616A.
- Safran, C. (1993, June). Dangerous obsession: The truth about repressed memories. *McCall's*, pp. 98-109, p. 155.
- Sandfort, T. G. M. (1992). The argument for adult-child sexual contact: A critical appraisal and new data. In W. O'Donohue & J. H. Geer (Eds.), *The sexual abuse of children: Theory and research* (pp. 38-48). Hillsdale, NJ: Lawrence Erlbaum Associates.

- Saunders, T. R. (1991). An overview of some psycholegal issues in child physical and sexual abuse. *Psychotherapy in Private Practice*, 9, 61-78.
- Seltzer, R., Lopes, G. M., Venuti, M. (1990). Juror ability to recognize the limitations of eyewitness identifications. *Forensic Reports*, 3, 121-137.
- Simon, J. L. (1992). *Resampling: The New Statistics*. Belmont, CA: Duxbury Press.
- Steller, M. (1989). Recent developments in statement analysis. In J. C. Yuille (Ed.), *Credibility assessment* (pp. 135-154), NATO ASI Series D: Behavioural and Social Sciences - Vol. 47. London: Kluwer Academic Publishers.
- Steller, M. (1992). Commentary: Rehabilitation of the child witness. In J. Doris (Ed.), *The suggestibility of children's recollections: Implications for eyewitness testimony* (pp. 106-109). Washington, DC: American Psychological Association.
- Steller, M., & Köhnken, G. (1989). Criteria-based statement analysis. In D. C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence* (pp. 217-245). New York: Springer.
- Steward, M. S., Bussey, K., Goodman, G. S., & Saywitz, K. J. (1993). Implications of developmental research for interviewing children. Special issue: Clinical recognition of sexually abused children. *Child Abuse and Neglect*, 17, 25-37.
- Tessler, M., & Nelson, K. (1994). Making memories: The influence of joint encoding on later recall by young children. Special issue: The recovered memory/false memory debate. *Consciousness and Cognition an International Journal*, 3, 307-326.
- Trankell, A. (1972). *Reliability of evidence*. Stockholm: Beckmans.
- Tully, B., & Tam, K. (1987). Helping the police with their inquiries: The development of special care questioning techniques. *Children & Society*, 3, 187-197.
- Tye, M. J. C. (1994). *Criteria-based content analysis of children's statements about a mock crime compared with the evaluations of naive subjects*. Unpublished master's thesis, University of North Dakota, Grand Forks.

- Tye, M. J. C., Honts, C. R., Amato-Henderson, S. (1994). [Lay subject evaluation performance under conditions of maximal information about scenario generating children's statements]. Unpublished raw data.
- Undeutsch, U. (1989). The development of statement reality analysis. In J. C. Yuille (Ed.), *Credibility assessment* (pp. 101-119), NATO ASI Series D: Behavioural and Social Sciences – Vol. 47. London: Kluwer Academic Publishers.
- Vogeltanz, N. D., & Drabman, R. S. (1995). A procedure for evaluating young children suspected of being sexually abused. *Behavior Therapy*, 26, 579-597.
- Vondergeest, L. K. (1992). *Expert testimony about statement validity assessment: Does it Usurp or enlighten?* Unpublished master's thesis, University of North Dakota, Grand Forks.
- Vrij, A., & van Wijngaarden, J. J. (1994). Will truth come out? Two studies about the detection of false statements expressed by children. *Expert Evidence: The International Digest of Human Behaviour Science and Law*, 3, 78-83.
- Wakefield, H., & Underwager, R. (1989). Evaluating the child witness in sexual abuse cases: Interview or inquisition? *American Journal of Forensic Psychology*, 7, 43-69.
- Wasserman, S., & Rosenfeld, A. (1992). An overview of the history of child sexual abuse and Sigmund Freud's contributions. In W. O'Donohue & J. H. Geer (Eds.), *The Sexual abuse of children: Theory and research* (pp. 49-72). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Wegener, H. (1989). The present state of statement analysis. In J. C. Yuille (Ed.), *Credibility assessment* (pp. 121-133), NATO ASI Series D: Behavioural and Social Sciences - Vol. 47. London: Kluwer Academic Publishers.
- Wells, G. L., & Loftus, E. F. (1992). Commentary: Is this child fabricating? Reactions to a new assessment technique. In J. Doris (Ed.), *The suggestibility of children's recollections: Implications for eyewitness testimony* (pp. 168-171). Washington, DC: American Psychological Association.

- Westcott, H. L., Davies, G. M., & Clifford, B. R. (1991). Adults' perceptions of children's videotaped truthful and deceptive statements. *Children & Society*, 5, 123-135.
- White, S. (1991). Mandated child abuse reporting. *Ethics and Behavior*, 1, 145-153.
- White, S. (1993). Assessing child sexual abuse allegations in custody disputes. In L. VandeCreek, S. Knapp, & T. L. Jackson (Eds.), *Innovations in clinical practice: A source book* (vol. 12., pp. 15-34). Sarasota, FL: Professional Resource Press/Professional Resource Exchange, Inc.