

Detecting Deception in Children: Event Familiarity Affects Criterion-Based Content Analysis Ratings

Kathy Pezdek, Anne Morrow, and
Iris Blandon-Gitlin
Claremont Graduate University

Gail S. Goodman
University of California, Davis

Jodi A. Quas
University of California, Irvine

Karen J. Saywitz
Harbor University of California, Los Angeles, Medical Center

Sue Bidrose
University of Otago

Margaret-Ellen Pipe
National Institute of Child Health and Human Development

Martha Rogers and Laura Brodie
Tustin, California

Statement Validity Assessment (SVA) is a comprehensive credibility assessment system, with the Criterion-Based Content Analysis (CBCA) as a core component. Worldwide, the CBCA is reported to be the most widely used veracity assessment instrument. We tested and confirmed the hypothesis that CBCA scores are affected by event familiarity; descriptions of familiar events are more likely to be judged true than are descriptions of unfamiliar events. CBCA scores were applied to transcripts of 114 children who recalled a routine medical procedure (control) or a traumatic medical procedure that they had experienced one time (relatively unfamiliar) or multiple times (relatively familiar). CBCA scores were higher for children in the relatively familiar than the relatively unfamiliar condition, and CBCA scores were significantly correlated with age. Results raise serious questions regarding the forensic suitability of the CBCA for assessing the veracity of children's accounts.

Between 1984 and 1990, America watched in horror while in Los Angeles, Raymond Buckey and his mother, Peggy McMartin Buckey, were charged with 65 counts of abuse, including rape, sodomy, fondling, oral copulation, and the drugging of children. The McMartin Preschool case made legal history for its sheer magnitude; the trial lasted 7 years, from its inception to the final

verdicts, and cost the state of California over \$16 million. In the end, the defendants were not convicted of any of the 65 counts against them. The McMartin trial shocked Los Angeles and the nation. One of the reasons was the disconcerting contrast between the fact that apparently, many of the child witnesses truly believed that what they reported did occur; yet the occurrence of many of the reported events seemed very unlikely. Sadly, the judicial and investigative procedures available were limited in their ability to determine which of the children's accounts described true events, if any, and which described false events, if any. The McMartin Preschool trial served as a call to action for researchers to study appropriate procedures for assessing the veracity of children's accounts of abusive events. We responded to this call and tested the validity of the Criterion-Based Content Analysis (CBCA; Raskin & Esplin, 1991b; Steller, 1989), reported by Vrij, Kneller, and Mann (2000) to be worldwide the most commonly used veracity assessment instrument.

Kathy Pezdek, Anne Morrow, and Iris Blandon-Gitlin, Department of Psychology, Claremont Graduate University; Gail S. Goodman, Department of Psychology, University of California, Davis; Jodi A. Quas, Program in Social Ecology, University of California, Irvine; Karen J. Saywitz, Department of Psychology and Biobehavioral Science, Harbor University of California, Los Angeles (UCLA), Medical Center; Sue Bidrose, Department of Psychiatry, University of Otago, Dunedin, New Zealand; Margaret-Ellen Pipe, Section on Social and Emotional Development, National Institute of Child Health and Human Development; Martha Rogers and Laura Brodie, independent practice, Tustin, California.

This article is based on work supported by the National Science Foundation to Kathy Pezdek under Grant No. SES-0111240. Any opinions, findings, and conclusions or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Correspondence concerning this article should be addressed to Kathy Pezdek, Department of Psychology, School of Behavioral and Organizational Sciences, Claremont Graduate State University, 123 East Eighth Street, Claremont, California 91711-3955. E-mail: kathy.pezdek@cgu.edu

It is particularly difficult to evaluate the veracity of children's accounts of abuse because usually (a) there is no physical evidence, and (b) the only witnesses to the crime are the victim and the perpetrator. The increase in allegations of child sexual abuse (Sedlak & Broadhurst, 1996) combined with the difficulty evaluating the validity of children's accounts (see, e.g., Ceci & Bruck, 1993) has increased the need for a valid and reliable validity assessment technique. The CBCA was developed in Germany in

the 1950s by Udo Undeutsch (see Undeutsch, 1989) and revised to its present form by Steller and Kohnken (1989) and Raskin and Esplin (1991b). The CBCA is the core component of the Statement Validity Assessment (SVA) system, a comprehensive system used for credibility assessment of sexual abuse allegations (see Raskin & Esplin, 1991b, for a detailed description of the SVA). The SVA includes three components: (a) a structured interview with the victim, (b) an assessment of a transcription of the victim's account using the CBCA, and (c) a validity checklist that incorporates information from the interview and the CBCA rating. Although Raskin and Esplin (1991a) emphasized the importance of using the entire SVA system when assessing allegations, the CBCA is considered the most important component of the SVA system by researchers (Steller, 1989; Ruby & Brigham, 1997), and accordingly, the CBCA has frequently been assessed for its stand-alone value (see Ruby & Brigham, 1997, for a review of this research).

The development of the CBCA was based on what Steller (1989) referred to as the Undeutsch hypothesis. According to the Undeutsch hypothesis, an account based on memory of a self-experienced event differs in content and quality from an account of an event that is imagined or suggested. From this notion, the 18 CBCA criteria were generated to discriminate between true accounts and fabricated accounts of alleged abuse, most commonly children's accounts of sexual abuse. These 18 criteria are assumed to be present in narrative accounts of events that actually did occur. The Appendix lists the 18 CBCA criteria and includes a brief description of each. These criteria are organized into three major categories.

According to Raskin and Esplin (1991a), the first category, General Characteristics of the Statement, includes the three most important criteria of the CBCA. This category is based on the premise that a report of a self-experienced event will be logical and coherent (Criterion 1), will have digressions or shifts in focus (Criterion 2), and will contain a wealth of details that relate to the people and acts in the event (Criterion 3). The other two categories, Specific Contents of the Statement and Motivation-Related Contents, refer to cognitive and motivational factors that are likely to be present in true statements. For instance, a truthful person is more likely to report verbatim content of speech or conversation (Criterion 6), and is more likely to doubt his or her own memory of events (Criterion 15). An account is considered likely to be true if a substantial number of the 18 criteria are present, with the first 3 criteria being necessary but not sufficient. Currently, however, no formalized system exists to weigh different criteria or to obtain absolute cut-off scores that can be used to designate an account as true or false.

Although the reliability of the CBCA has been convincingly demonstrated (Horowitz et al., 1997), support for the validity of the technique is less clear. Landry and Brigham (1992) compared judges trained to use the CBCA with untrained judges in their ability to discriminate between accounts of adults describing a true or a false personally traumatic event. Although the CBCA-trained judges were significantly more accurate classifying the accounts (55.3%) than were the untrained judges (46.9%), their level of performance was not impressive given the chance rate of 50%. Tye, Amato, Honts, Devitt, and Peters (1999) reported similar results with children. CBCA-trained judges correctly classified 89% of the accounts of children describing an observed mock theft that they were encouraged to describe correctly or to lie about. The

level of accuracy for CBCA judges compared favorably with that of untrained judges, and untrained judges, although significantly better than chance, achieved only 56% (Experiment 1) and 65% (Experiment 2) accuracy. Similar results have been reported in a number of other laboratory studies (Yuille, 1988; Zaparniuk, Yuille, & Taylor, 1995). However, although true events could be distinguished from false events in these studies using CBCA criteria, these two distributions were largely overlapping.

In the first reported field test of the CBCA in the United States, Esplin, Boychuk, and Raskin (1988) classified 20 child abuse cases as confirmed based on confessions by the perpetrator or strong physical evidence. Twenty different cases were classified as unconfirmed based on children's recantations, judicial dismissal, lack of prosecution, or no corroborating evidence. A CBCA-trained judge analyzed transcripts of children's reports in these cases. Significantly higher CBCA scores resulted for the confirmed than the unconfirmed reports; however, not all of the unconfirmed cases in this study were supported by independent evidence that the allegations were false.

More modest results were reported in two subsequent field studies. Craig, Scheibe, Raskin, Kircher, and Dodd (1999) reported that CBCA scores were significantly higher for children's accounts of confirmed than unconfirmed cases of abuse. Lamb et al. (1997) reported that although children's accounts of events rated as likely to have occurred received higher CBCA scores than those rated unlikely to have occurred, the two groups could only be differentiated on 6 of the 14 CBCA criteria used. In addition, Lamb et al. (1997) reported that the correlation between age of the child (range: 4–13) and CBCA scores was significant ($r = .40$). A significant effect of age on CBCA scores was also reported by Vrij, Akehurst, Soukara, and Bull (2002), with older children, ages 10–15, compared with undergraduate students.

In a similar study, Lamers-Winkelmann, Buffing, and van der Zanden (1992) examined the accounts of 103 children who alleged sexual abuse. Children's claims were independently classified as substantiated, highly probable, or unfounded. Judges using the CBCA criteria successfully discriminated among these three groups with older children but not younger children. Concern about the appropriateness of the CBCA for young children was bolstered by the report of Lamers-Winkelmann and Buffing (1996) that for children 2–12 years of age, with decreasing age, the number of the CBCA criteria not fulfilled increased. This suggests that all of the CBCA criteria may not be appropriate with younger children. Despite the widespread use of the CBCA throughout the United States and Europe, in a review of the research on the validity of the CBCA with children, Lamb et al. (1997) recently concluded that the CBCA is currently not sufficiently precise to be suitable for forensic application.

Using an experimental paradigm to assess the validity of the CBCA, Steller, Wellershaus, and Wolf (1988; as cited in Steller, 1989) asked individual children to provide a true and a fabricated account of an upsetting event that they had personally experienced. Three CBCA-trained judges rated each account. Several subjects described medically related upsetting events. For accounts of these events, 11 CBCA criteria were strongly present in the true accounts but not in the fabricated accounts. However, for the accounts of more everyday events (e.g., fight with another child), true and fabricated accounts received similar ratings from the judges. Thus, the CBCA more effectively discriminated between

true and false accounts of unfamiliar events than between true and false accounts of familiar events.

Together, these results suggest that although the CBCA does discriminate between true and fabricated accounts of events, the discriminability of the CBCA is not consistently impressive, and there are several factors that may systematically bias CBCA scores. The role of these factors may ultimately undermine the suitability of the CBCA for forensic applications. These factors include the child's age and cognitive abilities and perhaps the familiarity of the event. The purpose of this study is to test the specific hypothesis that accounts of familiar events, regardless of truth value, receive higher CBCA scores than do accounts of unfamiliar events. In accordance, because many of the cases in which children are relied on as witnesses are cases of abuse, and abuse involves behaviors that are typically unfamiliar to children, based on CBCA ratings their accounts of such events would often be erroneously classified as false.

Pezdek, Finger, and Hodge (1997) and Pezdek and Hodge (1999) demonstrated that the plausibility of an event can affect people's memory for the event. In their experiments, they reported that it was easier to plant false memories of plausible events than implausible events. Adults and children were tested on their memories of several true events and two target false events. The two false target events, each alleged to have occurred at a young age, were being lost in a shopping mall (plausible) and receiving an enema (implausible) at a young age. Significantly more participants reported falsely remembering the plausible event as opposed to the implausible event as having occurred. The interpretation of the finding that the plausible event was more likely to be judged as true than the implausible event was based on the assumption that more script-relevant knowledge exists in memory for plausible than for implausible events. When a suggested false event is judged to be true, details of the generic script for the event, as well as details from related episodes of the event, are "transported" to the memory for the suggested false event. As a consequence, memory for the false event becomes developed by this related information in memory. In accordance, memories for false plausible events take on the appearance of memories for true events and are more likely to be reported as such. Extrapolating this finding to the veracity assessment literature, it is predicted that memory for episodes of familiar events will contain more script-relevant information than will memory for episodes of unfamiliar events. As a consequence, CBCA scores of accounts of familiar events are likely to appear true, and CBCA scores of accounts of unfamiliar events are likely to appear false. The present study tests this hypothesis.

In this study, CBCA scores were obtained from transcripts of children's description of a stressful and invasive medical procedure called voiding cystourethrogram fluoroscopy (VCUG). This procedure is used to identify reflux, which can be the cause of urinary tract dysfunction in children. The VCUG procedure is painful and involves intrusive, forced genital contact (catheterization through the urethra and infusion of a liquid into the bladder). Even the doctors administering the procedure admit that in many ways the VCUG procedure is similar to sexual assault on a child. The VCUG is not an elective procedure for the children. Each child had received a VCUG one time (the relatively unfamiliar condition) or multiple times (the relatively familiar condition). In the control condition, a different group of children described an

anal-genital exam that was part of a routine medical visit. Transcripts of children's accounts of these procedures were then coded by two CBCA-trained coders. The VCUG procedure was used as the target event in this study because it is similar in many ways to child sexual abuse, the real world behavior that we hope to generalize these results to. However, whereas we knew exactly what had occurred in the medical procedure that each child in this study participated in (the procedures were videotaped), practitioners rarely know what transpired in alleged incidents of sexual abuse.

Method

Pilot Test of the Similarity Between the VCUG Procedure and Child Sexual Abuse

In this study, the VCUG procedure was used as a proxy for sexual abuse. However, it is important to demonstrate that the VCUG procedure and child sexual abuse are similar if the results of this study are to be generalized to real-world incidents of child sexual abuse. A pilot study was conducted to determine the similarity between the VCUG procedure and child sexual abuse. In this pilot study, first a list was generated that included 14 features common to sexual abuse of children between the ages of 4 and 12. This list was then distributed to 24 counselors who work with abused children at The Family Stress Center of the Child and Family Guidance Center in Los Angeles. The mean number of years that this group of 24 counselors had worked with abused children was 6.10 years. Each of these counselors reviewed the list of 14 features and rated each on a scale, 1 = *almost always a feature of child sexual abuse*, 2 = *frequently a feature of child sexual abuse*, 3 = *occasionally a feature of child sexual abuse*, and 4 = *rarely a feature of child sexual abuse*. The counselors were then asked to add to the list any additional features that they thought were common components of child sexual abuse and were asked to rate each on the same 1–4 scale indicated above.

The responses from the counselors were tallied, and a new list was generated that included all features that were rated 1 or 2 by at least 70% of the counselors. This second list included eight features. This list can be thought of as the defining features of child sexual abuse for the 24 counselors tested. The next step was to see if these eight features were considered common features of the VCUG procedure, as well, by practitioners who frequently administer VCUGs to children in this same age range.

The eight-item list was distributed to 11 practitioners who administer VCUGs to children at Children's Hospital of Los Angeles. These included six radiologists and five X-ray technicians. The mean number of years that these 11 practitioners had administered VCUGs to children was 16.73 years. The total number of VCUGs administered to children by each ranged from "hundreds" to "about 3,000." These individuals reviewed the list of eight features and rated each on the same 1–4 scale indicated above but in terms of how common each feature was to the VCUG procedure. This list of eight features, with the mean rating for each by the VCUG practitioners, is included in Table 1.

The majority of the features that defined child sexual abuse for the counselors were also considered to be at least *frequently* associated with the VCUG procedure. Five of the eight features received mean responses of 2.00 or less, indicating that these features were between *almost always* and *frequently* a feature of the VCUG procedure. In addition, two of the eight features received mean responses of 2.54, indicating that these features were between *frequently* and *occasionally* features of the VCUG procedure. Only one feature, "Child tries to distract him/herself during incident," received a rating of 3.00, indicating that the feature is on average only *occasionally* a feature of the VCUG procedure. These results suggest a high rate of similarity between the features common to child sexual abuse

Table 1
Features Rated as Most Common to Child Sexual Abuse

Features of child sexual abuse	Mean rating by VCUG practitioner
1. Genital contact	1.20
2. Contact forced against will of child	2.54
3. Child's clothes are removed	1.55
4. Child is anxious	2.00
5. Child is encouraged to relax and not resist	1.00
6. Child tries to distract him/herself during incident	3.00
7. Entire incident is not voluntary	2.54
8. Incident carried out by trusted authority figure	1.09

Note. Ratings were based on a scale ranging from 1 (*almost always a feature of the VCUG procedure*) to 4 (*rarely a feature of the VCUG procedure*). VCUG = voiding cystourethrogram fluoroscopy.

and those common to the VCUG procedure and thus endorse the generalization of the findings from this study to real world incidents of child sexual abuse.

Participants and Design

The sample in this study consisted of 114 children (86 girls and 28 boys) who had participated in one of three previous studies examining children's memory for a traumatic event or a medical procedure. These include studies by Goodman, Quas, Batterman-Faunce, Riddlesberger, and Kuhn (1994), Quas et al. (1999), and Saywitz, Goodman, Nicholas, and Moan (1991). From these previous studies, three groups of children were selected to be included in the study: (a) children receiving the VCUG procedure for the first time (the relatively unfamiliar condition, $n = 49$), (b) children receiving the VCUG procedure who had previously had the procedure two or more times (the relatively familiar condition, $n = 27$), (c) children receiving an anal-genital exam for medical reasons, but who had never received the VCUG procedure (the control condition, $n = 38$). An anal-genital exam was used as the control condition because this procedure is similar to the VCUG in that it involves contact with a private part of the body and is also likely to involve some amount of discomfort and/or embarrassment.

The age range of the children was from 3 to 13.6 years ($M = 6.51$, $SD = 1.2$). The mean age of participants in the relatively familiar condition ($M = 6.6$ years), the relatively unfamiliar condition ($M = 6.2$ years), and the control condition ($M = 6.8$ years) did not significantly differ, $F(2, 111) = 1.17$. The ethnic composition of the sample was as follows: 78.5% Caucasian, 8.4% other-mixed racial, 5.6% African American, 3.7% Hispanic, 2.8% Asian, and .9% Middle Eastern.

Procedure

Forty-nine of the children (17 in the relatively familiar condition, 32 in the relatively unfamiliar condition) were selected from the Goodman et al. (1994) study in which children's memory for the VCUG exam was investigated. In this study, parents of children scheduled to undergo a VCUG were contacted and each agreed to allow his or her child to participate in a study of memory for the procedure. A research assistant met each child and parent(s) in the waiting room of the hospital radiology department on the day of the VCUG exam to observe and videotape the procedure. After a delay of 1–2 weeks ($M = 1.44$ weeks), each child returned to the university laboratory for a memory interview. All interviewers were trained to follow a uniform interview sequence. The memory questions began with a free-recall question concerning the child's memory for what happened during the VCUG exam. Several prompts were then given to focus the

child on talking about the medical procedure. The transcript of this portion of the interview served as part of the data in this study.

The second set of participants ($n = 27$; 10 in the relatively familiar condition, 17 in the relatively unfamiliar condition) were selected from the Quas et al. (1999) study examining children's memory for the VCUG exam. In this study, parents of children who had undergone a VCUG in the past were contacted. The delay between when the child had undergone the target VCUG exam and the memory interview ranged from 4 to 144 weeks ($M = 28.4$ weeks). All interviewers were trained to follow a uniform interview sequence. A research associate began the child interview with free-recall questions concerning the child's memory for what happened during the VCUG exam. The transcript of this portion of the interview served as part of the data in this study.

The participants in the control condition ($N = 38$) were selected from the Saywitz et al. (1991) study examining children's memory for a routine physical examination involving genital touch. The study was conducted at a medical center where the routine physical checkup was administered. The checkup also included an anal-genital examination in which the exterior of these areas was visually inspected and touched to look for rashes, infections, torn tissue, and other evidence of trauma. Each child returned with a parent 1 to 4 weeks later ($M = 2.64$ weeks) for the memory interview. All interviewers were trained to follow a uniform interview sequence. The interview began with free-recall questions concerning the child's memory for what happened during their previous visit to the doctor's office. The transcript of this portion of the interview served as part of the data in the control condition.

The mean time delay between when each child experienced the target medical procedure and when the child was interviewed did not significantly differ among the three conditions, $F(2, 111) = 2.68$, nor was the difference in time delay significant between the relatively familiar and unfamiliar conditions (mean familiar condition = 10.08 weeks; mean unfamiliar condition = 11.6 weeks), $t(74) < 1.00$.

CBCA Rating of the Data

The memory interview of each child was transcribed verbatim, and typed copies were provided to two CBCA-trained judges. These two judges had no other role in this study except to rate the transcripts. Both raters are clinical forensic psychologists who attended a 3-day training seminar of David Raskin. Both raters also have extensive experience with the CBCA both in their research and clinical practice. One of the raters received additional training from David Raskin and served as a rater in his research. She was also supervised by him in several legal cases involving forensic application of the CBCA. The other rater used the CBCA in her dissertation research; David Raskin and the other rater in this study were members of her dissertation committee. Judges were unaware of the purpose of the study and the specific experimental conditions.

Prior to coding the transcripts, the two judges viewed a medical videotape showing the VCUG procedure. A description of the VCUG procedure was also read aloud to each judge as follows:

For the VCUG test, the child first changed into a hospital gown. Then the child and the parent were brought into an X-ray room by a technician for initial X-rays. Once the child was on the examination table, and the X-ray camera was in place, the parent, RA, and technician stood at the doorway outside the room while the X-ray was taken. Next, a nurse and technician prepared the child to be catheterized by washing the child's genitals with soap. For girls, this entailed bringing their legs up to a "frog-like" position while lying on their backs. After the cleansing, some boys were given the benefit of an anesthetizing gel inserted into the tip of the penis to numb it slightly; however, the gel was only partially effective. Next, most children were given a pinwheel to blow on; this served both to distract them and to facilitate relaxation of the stomach muscles. The technician and parent held the child down and tried to comfort him or her while the

nurse performed the catheterization (insertion of a 12" to 15" plastic tube up through the urethra into the middle of the child's bladder). If a child struggled excessively, extra technicians helped restrain the child or the child was physically restrained under special equipment (e.g., tied onto a "papoose board"). After the catheter was inserted, the parent was required to leave the room before the doctor entered. A liquid dye (contrast medium) was then infused through the catheter into the child's bladder as the doctor took X-rays. Once the child's bladder was full, he or she was asked to void on the examination table. X-rays were taken as the child voided; the catheter slipped out nonpainfully at this time. When the child would not or could not urinate on the table, a nearby water faucet was turned on or warm water was poured over their genitals. (From "Predictors of accurate and inaccurate memories of traumatic events experienced in childhood," by G. S. Goodman, J. A. Quas, J. M. Batterman-Faunce, M. M. Riddlesberger, & J. Kuhn, 1994, *Consciousness and Cognition*, 3, p. 278. Copyright 1994 by Elsevier. Reprinted with permission.)

The judges were then told that some of the children received a VCUg and some of the children received a different (but undisclosed) medical procedure. The judges were instructed that they would be coding the transcripts because they had been trained to use the CBCA to help determine the veracity of the events described. The judges were provided with a coding sheet that included a list of 16 CBCA criteria and a brief definition of each. Criteria 17 (Self-Depreciation) and 18 (Pardoning the Accused) were not used for coding because they are not applicable to the target event.¹ Items with reference to sexual events were rewritten to be appropriate for this study.

Following the procedure of Boychuk (1991), Landry and Brigham (1992), Steller (1989), and Tye, Henderson, and Honts (1995), each of the 16 criteria received a rating of either 2 (*criterion strongly present*), 1 (*criterion present*), or 0 (*criterion absent*). Each judge's score was the sum of all the criteria present in each transcript (maximum total score = 32). The correlation between the ratings for the two judges across all subjects was $r = .97$, $p < .01$. Because this correlation was significant, a mean CBCA rating for each transcript was computed based on the two judges' ratings. An item-by-item analysis using Cohen's kappa revealed that interjudge agreement on each of the 16 criteria ranged from .81 to 1.0, with a mean agreement of .92.

Verbal Familiarity with VCUg Procedure

A measurement of the children's verbal familiarity with the VCUg exam was included to assess whether children who discussed the exam with their parents produced higher CBCA scores than did children who did not discuss the exam with their parents. It was hypothesized that children who discussed the exam with their parents would be more familiar with the VCUg procedure and would consequently receive higher CBCA scores. Each participant from the study by Goodman et al. (1994; $N = 49$) had been questioned regarding how much he or she discussed the VCUg procedure with his or her parent. In addition, each mother was asked to provide a "yes" or "no" response to the following written statements: "Discussed exam with child," "Explained exam to child," "Child has talked about exam," "Child has wanted to talk about exam," "Child has asked questions about exam," and "Child has sought explanation about exam." A score was computed for each question (0 = no, 1 = yes), and a composite verbal familiarity score was computed as the number of "yes" responses to all 6 questions (possible range of 0 to 6 points).

Results

Analyses of CBCA Scores

The mean CBCA score for each of the conditions is presented in Table 2. The data were first analyzed to assess CBCA scores in the

Table 2
Mean CBCA scores (and Standard Deviations) for Each Condition

Condition	Mean CBCA score
Relatively familiar ($n = 27$)	10.67 (5.73)
Relatively unfamiliar ($n = 49$)	7.68 (5.97)
Control ($n = 38$)	9.92 (5.07)

Note. CBCA = Criterion-Based Content Analysis.

two VCUg conditions because they compare to the control condition. A one-way analysis of variance indicated a significant difference in CBCA scores across the three conditions, $F(2, 111) = 3.16$, $p < .05$, $r^2 = .05$. Planned comparisons were then conducted. As predicted, children in the relatively familiar condition, $M = 10.67$, received significantly higher CBCA scores than did children in the relatively unfamiliar condition, $M = 7.68$, $F(1, 74) = 4.69$, $p < .05$, $r^2 = .06$. Although the difference in mean CBCA scores between children in the relatively familiar condition, $M = 10.67$, and children in the control condition, $M = 9.92$, was not significant, $F(1, 63) < 1.00$, $r^2 = .004$, the difference between children in the relatively unfamiliar condition, $M = 7.68$, and children in the control condition was marginally significant, $M = 9.92$, $F(1, 85) = 4.00$, $p = .058$, $r^2 = .04$. The high CBCA scores for children in the control condition were not surprising given that these children described a routine medical examination that did, in fact, occur.

CBCA Scores and Verbal Familiarity With VCUg Procedure

Analyses were conducted to explore the relationship between verbal familiarity with the VCUg exam and CBCA scores. Verbal familiarity was assessed by several questions in which mothers of the 49 children who participated in the study by Goodman et al. (1994) were asked to indicate how much their children discussed the exam. There was a significant correlation between CBCA scores and the composite verbal familiarity scores, $r = .34$, $p < .05$. Children whose mothers had indicated that they discussed the VCUg exam with them received higher CBCA scores than did children whose mothers had indicated that they did not discuss the VCUg exam. The number of VCUg exams received was not significantly correlated with overall verbal familiarity score, $r = .10$, $p = .50$, nor was the correlation significant between children's age and the verbal familiarity score, $r = .16$, $p = .26$.

CBCA Score and Age

The next analysis was conducted to explore the relationship between children's age at the time they were interviewed (range:

¹ After reviewing a subset of our transcripts, Phillip Esplin suggested that all of the original 18 CBCA criteria be used with the exception of Criteria 17 and 18. Although Kohnken, Schimossek, Aschermann, and Hofer (1995) tested an additional five criteria for inclusion in the CBCA, their own support for including these additional criteria is weak. From their research they concluded, "Does this mean that the additional content characteristics should be included in the list of CBCA criteria? Such a step would certainly be premature, and we would not recommend it" (p. 681).

3–13.58 years, $M = 6.51$) and CBCA scores. Averaged across all participants, there was a significant correlation between children's age and CBCA scores, $r = .20$, $p < .05$.

Differences Between the Two VCUG Groups on Each of the Three Categories of CBCA Criteria

Using a multivariate analysis of variance (MANOVA), an additional analysis was conducted to assess whether there were significant differences between the relatively familiar and unfamiliar VCUG groups on each of the three categories in the CBCA. These included General Characteristics (Criteria 1–3), Specific Contents (Criteria 4–13), and Motivation-Related Contents (Criteria 14–16). Within each of these three categories, the judges' ratings for each of the criteria were summed. These three sums served as the three dependent measures in the MANOVA.²

Prior to conducting the MANOVA, the Mahalanobis distance procedure suggested by Tabachnick and Fidell (2001) was used to identify multivariate outliers. Two multivariate outliers were identified and removed from the relatively unfamiliar condition. The overall MANOVA, which included the three categories of criteria, did not reach statistical significance, $F(3, 70) = 2.09$, $p < .11$, $\eta^2 = .08$. An inspection of the data revealed that in the third category of CBCA criteria, Motivation-Related Content, the large majority of participants in both groups received a score of 0 on each of the three criteria. Consequently, there was little difference between the groups to be accounted for. Thus, a second MANOVA was conducted that included only the first two CBCA categories and not the Motivation-Related Content category. This MANOVA was significant overall, $F(2, 71) = 3.18$, $p < .05$, $\eta^2 = .08$, as well as for the General Characteristics category, $F(1, 72) = 3.98$, $p < .05$, $\eta^2 = .05$, and the Specific Content category, $F(1, 72) = 6.43$, $p < .02$, $\eta^2 = .08$, separately. Thus, with the Motivation-Related Content category removed, the relatively familiar and unfamiliar groups could be differentiated on the basis of both the General Characteristics category and the Specific Content category of the CBCA.

Discussion

The purpose of this study was to test if CBCA scores were affected by individuals' familiarity with the events described by using a unique and traumatic target event. The major finding was that CBCA scores were significantly higher for transcripts of children's accounts of having received a VCUG in the familiar condition than in the unfamiliar condition; children who had received a VCUG more than one time received higher CBCA scores than did children who had received a VCUG only once. Thus, CBCA scores are affected by the familiarity of the event being described; accounts of familiar events are more likely to be considered true than are accounts of unfamiliar events.

This finding is also supported by the result that children who talked more about the VCUG procedure with their parent also produced higher CBCA scores. Mothers in the subset of participants from the study by Goodman et al. (1994) had been questioned regarding how much they talked with their children about the VCUG procedure. A composite score of each child's verbal familiarity with the VCUG was significantly correlated with CBCA scores for these children; children whose mothers had

indicated that they discussed the VCUG exam with them received higher CBCA scores than did children whose mothers had indicated that they did not discuss the VCUG exam with them. These findings suggest that CBCA scores are affected by the familiarity of the event being described and that familiarity can result from repeated experience with the event as well as from repeatedly talking about the experienced event.

In light of the easy access to sexually explicit material in our culture and the commonness of conversations about sexual topics, many children have knowledge of sexual behavior that was acquired indirectly from these sources. A child's description of an alleged incident of sexual abuse may include information about the alleged incident as well as information acquired indirectly from other sources. To the extent that a child is familiar with the alleged incident, their account of the incident is likely to have characteristics that produce a high CBCA score, indicating that the event did occur whether it did in fact occur or not.

Why do familiar events receive higher CBCA scores than unfamiliar events? The experience of a specific event is not stored as an isolated unit in memory. Rather, when a specific event is encoded into memory, related script-relevant knowledge is activated that aids in the comprehension of the event. This script-relevant knowledge includes generic information about the event acquired from prior knowledge and experience and is stored in memory along with information retained from experiencing the event directly. By definition, individuals have more script-relevant knowledge in memory for familiar events than for unfamiliar events. Thus, children's accounts of familiar events are likely to contain more information and have a more coherent structure than are their descriptions of unfamiliar events. As a consequence, CBCA scores would be higher for familiar events than for unfamiliar events.

Several researchers have expressed concerns about the effects of age and developmental level on CBCA scores (Lamers-Winkelman et al., 1992; Lamers-Winkelman & Buffing, 1996; Lamb et al., 1997; Raskin & Esplin, 1991a). The results of the present study provide additional support for this concern. There was a significant correlation between age and CBCA scores ($r = .20$). This is not surprising given that the General Characteristics of the CBCA are the most discriminating factors (i.e., the logical structure and amount of detail in the account), and narrative accounts of young children are less well-structured than are those of older children (Fivush & Slackman, 1986; Fivush, Kuebli, & Clubb, 1992).

Together these results suggest serious concerns about the forensic suitability of the CBCA for discriminating between children's accounts of true and false accounts of events. CBCA scores are

² The complete MANOVA analyses and table are available from Kathy Pezdek on request. The MANOVA using all 16 CBCA criteria to predict differences between the relatively familiar and unfamiliar groups was not significant, $F(15, 16) < 1.00$. This is because on many of the CBCA criteria, the large majority of participants receive scores of 0. Lamb et al. (1997) and others have reported similar findings. Thus, there was no difference between the groups to be accounted for. This limitation was avoided in the MANOVAs reported in this article by summing across all CBCA criteria in each of the three CBCA categories and by using these sums, rather than the scores for the 16 individual criteria, as the predictor variables.

significantly affected by both the familiarity of the event described and the age of the child providing the description.

References

- Boyчук, T. D. (1991). *Criteria-based content analysis of children's statements about sexual abuse: A field-based validation study*. Unpublished doctoral dissertation, Arizona State University, Tempe.
- Ceci, S. J., & Bruck, M. (1993). The suggestibility of the child witness: A historical review and synthesis. *Psychological Bulletin*, *113*, 403–439.
- Craig, R. A., Scheibe, R., Raskin, D. C., Kircher, J. C., & Dodd, D. (1999). Interviewer questions and content analysis of children's statements of sexual abuse. *Applied Developmental Science*, *3*, 77–85.
- Esplin, P. W., Boyчук, T., & Raskin, D. C. (1988, June). *A field study of criteria-based content analysis of children's statement in sexual abuse cases*. Paper presented at the NATO Advanced Studies Institute on Credibility Assessment, Maratea, Italy.
- Fivush, R., Kuebli, J., & Clubb, P. A. (1992). The structure of events and event representations: A developmental analysis. *Child Development*, *63*, 188–201.
- Fivush, R., & Slackman, E. (1986). The acquisition and development of scripts. In K. Nelson (Ed.), *Event knowledge: Structure and function in development* (pp. 71–96). Hillsdale, NJ: Erlbaum.
- Goodman, G. S., Quas, J. A., Batterman-Faunce, J. M., Riddlesberger, M. M., & Kuhn, J. (1994). Predictors of accurate and inaccurate memories of traumatic events experienced in childhood. *Consciousness and Cognition*, *3*, 269–294.
- Horowitz, S. W., Lamb, M. E., Esplin, P. W., Boyчук, T. D., Krispin, O., & Reiter-Lavery, L. (1997). Reliability of criteria-based content analysis of child witness statements. *Legal and Criminological Psychology*, *2*, 11–21.
- Kohnken, G., Schimossek, E., Aschermann, E., & Hofer, E. (1995). The cognitive interview and the assessment of the credibility of adults' statements. *Journal of Applied Psychology*, *80*, 671–684.
- Lamb, M. E., Sternberg, K. J., Esplin, P. W., Hershkowitz, I., Orbach, Y., & Hovav, M. (1997). Criteria-Based Content Analysis: A field validation study. *Child Abuse & Neglect*, *21*, 255–264.
- Lamers-Winkelmann, F., & Buffing, F. (1996). Children's testimony in the Netherlands: A study of statement validity analysis. *Criminal Justice and Behavior*, *23*, 304–321.
- Lamers-Winkelmann, F., Buffing, F., & van der Zanden, A. P. (1992, September). *Statement validity analysis in child sexual abuse cases: A field study*. Poster presented at the Third Conference of Law and Psychology, Oxford, England.
- Landry, K. L., & Brigham, J. C. (1992). The effect of training in criteria-content analysis on the ability to detect deception in adults. *Law and Human Behavior*, *16*, 663–676.
- Pezdek, K., Finger, K., & Hodge, D. (1997). Planting false childhood memories: The role of event plausibility. *Psychological Science*, *8*, 437–441.
- Pezdek, K., & Hodge, D. (1999). Planting false childhood memories in children: The role of event plausibility. *Child Development*, *70*, 887–895.
- Quas, J., Goodman, G., Bidrose, S., Pipe, M., Craw, S., & Ablin, D. (1999). Emotion and memory: Children's long-term remembering, forgetting, and suggestibility. *Journal of Experimental Child Psychology*, *72*, 235–270.
- Raskin, D. C., & Esplin, P. W. (1991a). Assessment of children's statements of sexual abuse. In J. Doris (Ed.), *The suggestibility of children's recollections* (pp. 153–164). Washington, DC: American Psychological Association.
- Raskin, D. C., & Esplin, P. W. (1991b). Statement validity assessment: Interview procedures and content analysis of children's statements of sexual abuse. *Behavioral Assessment*, *12*, 265–291.
- Ruby, C. L., & Brigham, J. C. (1997). The usefulness of the criterion-based content analysis technique in distinguishing between truthful and fabricated allegations: A critical review. *Psychology, Public Policy, and Law*, *3*, 705–737.
- Saywitz, K. J., Goodman, G. S., Nicholas, E., & Moan, S. F. (1991). Children's memories of a physical examination involving genital touch: Implications for reports of child sexual abuse. *Journal of Consulting and Clinical Psychology*, *59*, 682–691.
- Sedlak, A. J., & Broadhurst, D. D. (1996). *Third national incidence study of child abuse and neglect: Final report*. Washington, DC: U.S. Department of Health and Human Resources, Administration for Children and Families.
- Steller, M. (1989). Recent developments in statement analysis. In J. C. Yuille (Ed.), *Credibility assessment* (pp. 135–154). Dordrecht, Netherlands: Kluwer Academic Publishers.
- Steller, M., & Kohnken, G. (1989). Statement analysis: Credibility assessment of children's testimonies in sexual abuse cases. In D. C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence* (pp. 217–245). New York: Springer.
- Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston: Allyn and Bacon.
- Tye, M. J. C., Amato, S. L., Honts, C. R., Devitt, M. K., & Peters, D. (1999). The willingness of children to lie and the assessment of credibility in an ecologically relevant laboratory setting. *Applied Developmental Science*, *3*, 92–109.
- Tye, M. J. C., Henderson, S. A., & Honts, C. R. (1995, January). *Evaluating children's testimonies: CBCA and lay subjects*. Paper presented at CRIMECON: International INTERNET conference on Crime and Criminal Justice.
- Undeutsch, U. (1989). The development of statement reality analysis. In J. C. Yuille (Ed.), *Credibility assessment* (pp. 101–121). Dordrecht, the Netherlands: Kluwer Academic Publishers.
- Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2002). Will the truth come out? The effect of deception, age, status, coaching, and social skills on CBCA scores. *Law and Human Behavior*, *26*, 261–283.
- Vrij, A., Kneller, W., & Mann, S. (2000). The effect of informing liars about Criteria-Based Content Analysis on their ability to deceive CBCA-raters. *Legal and Criminological Psychology*, *5*, 57–70.
- Yuille, J. C. (1988). The systematic assessment of children's testimony. *Canadian Psychology*, *29*, 247–262.
- Zaparniuk, J., Yuille, J. C., & Taylor, S. (1995). Assessing the credibility of true and false statements. *International Journal of Law Psychiatry*, *18*, 343–352.

(Appendix follows)

Appendix

Content Criteria for Statement Analysis

General Characteristics

1. Logical Structure: Is the statement coherent? 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 be locally consistent.)
3. Quantity of Details: Are there specific descriptions of place and time? Are persons, objects, and events specifically described? (Note: Repetitions do not count.)

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. Interactions: Are there reports of actions and reactions or conversation composed of a minimum of three elements involving at least the accused and the witness?
6. Reproduction of speech: Is speech and conversation during the incident reported in its original form? (Note: Unfamiliar terms or quotes are especially strong indicators, even when attributed to only one participant.)
7. Unexpected Complications: Was there an unplanned interruption or an unexpected complication or difficulty during the sexual incident?
8. Unusual Details: Are there details of person, objects, or events that are unusual, yet meaningful in this context? (Note: Unusual details must be realistic.)
9. Superfluous Details: Are peripheral details described in connection with the alleged sexual events that are not essential and do not contribute directly to the specific allegations? (Note: If passage satisfies any of the specific criteria 4–18, it probably is not superfluous.)
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 a sexually toned event or conversation of a sexual nature that it is related in some way to the incident but is not part of the alleged offences?
12. Subjective Experience: Did the child describe feelings or thoughts experienced at the time of the incident? (Note: This criterion is not satisfied when the child responds to a direct question, unless the answer goes beyond the question.)
13. Attribution of the Accused's Mental State: Is there reference to the perpetrator's feelings or thoughts during the incident? (Note: Descriptions of overt behavior do not qualify.)

Motivation-Related Contents

14. Spontaneous Corrections or Additions: 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 or Knowledge: Did the child indicate lack of memory or knowledge of an aspect of the incident? (Note: In response to a direct question, the answer must go beyond "I don't know" or "I can't remember.")
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-Depreciation: Did the child describe some aspect of his/her behavior related to the sexual incident as wrong or inappropriate?
18. Pardoning the Accused: Did the child make excuses for or fail to blame the alleged perpetrator, or minimize the seriousness of the acts, or fail to add to the allegations when opportunity occurred?

Note. From "Statement Validity Assessment: Interview Procedures and Content Analysis of Children's Statements of Sexual Abuse," by D. C. Raskin and P. W. Esplin, 1991, *Behavioral Assessment*, 12, p. 279. Copyright 1991 by Elsevier. Reprinted with permission.

Received July 3, 2002

Revision received March 6, 2003

Accepted May 12, 2003 ■

Wanted: Your Old Issues!

As APA continues its efforts to digitize journal issues for the PsycARTICLES database, we are finding that older issues are increasingly unavailable in our inventory. We are turning to our long-time subscribers for assistance. If you would like to donate any back issues toward this effort (preceding 1982), please get in touch with us at journals@apa.org and specify the journal titles, volumes, and issue numbers that you would like us to take off your hands.