APPLIED COGNITIVE PSYCHOLOGY
Appl. Cognit. Psychol. **23**: 901–917 (2009)
Published online 11 August 2008 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/acp.1504

Criteria-based Content Analysis of True and Suggested Accounts of Events

IRIS BLANDÓN-GITLIN[1]*, KATHY PEZDEK[2], D. STEPHEN LINDSAY[3] and LISA HAGEN[4]

[1]*California State University, Fullerton, USA*
[2]*Claremont Graduate University, USA*
[3]*University of Victoria, Canada*
[4]*Carleton University, Canada*

SUMMARY

Worldwide, the criteria-based content analysis (CBCA) is probably the most widely used veracity assessment technique for discriminating between accounts of true and fabricated events. In this study, two experiments examined the effectiveness of the CBCA for discriminating between accounts of true events and suggested events believed to be true. In Experiment 1, CBCA-trained judges evaluated participants' accounts of true and suggestively planted childhood events. In Experiment 2, judges analysed accounts of recent events that were experimentally manipulated to be a (a) true experience, (b) false experience believed to be true and (c) deliberately fabricated experience. In both experiments CBCA scores were significantly higher for accounts of true events than suggested events. However, this difference was not significant for participants classified as experiencing 'full' memories for the suggested event. Self-report memory measures supported the findings of the CBCA analyses. Taken together these results suggest that the CBCA discriminative power is greatly constrained. Copyright © 2008 John Wiley & Sons, Ltd.

In forensic settings, it is difficult to assess the authenticity of sexual abuse allegations. This difficulty stems from the fact that the only witnesses are usually the victim and alleged perpetrator, and other corroborating evidence is often lacking (Raskin & Esplin, 1991b). A commonly used veracity assessment technique in forensic settings is the statement validity assessment (SVA) (see Raskin & Esplin, 1991a and Vrij, 2005 for a complete review of SVA). The core component and the most researched aspect of the SVA system is the criteria-based content analysis (CBCA). The CBCA procedure involves the systematic assessment of truthfulness in witnesses' spoken accounts of events. Empirical research has demonstrated that under some conditions the CBCA can be useful for differentiating between truthful and deceptive accounts. Almost all of this research has focused on evaluating the validity of the CBCA for differentiating between accounts of events that are known to be true and those that are known to be deliberately fabricated. One aspect of the CBCA that has not received systematic evaluation is its potential for differentiating between accounts of true events and accounts of false events reported without the intention

*Correspondence to: Iris Blandón-Gitlin, Department of Psychology, California State University, Fullerton, USA.
E-mail: iblandon-gitlin@fullerton.edu

to deceive. That is, accounts of false events that are subjectively true. The purpose of the present study is to address this question by testing the effectiveness of the CBCA for discriminating between true and suggested accounts of events.

The CBCA was originally developed over 50 years ago for the purpose of evaluating allegations of child witnesses in contested cases of sexual abuse. Although the CBCA was developed specifically for this purpose it can detect deception in accounts of adult participants (Kohnken, Schimossek, Aschermann, & Hofer, 1995; Landry & Brigham, 1992; Sporer, 1997) and has been useful with a range of events (Steller, 1989; Tye, Amato, Honts, Devitt, & Peters, 1999; Yuille, 1988). The development of the CBCA was based on what has been termed the Undeutsch hypothesis, which states that an account derived from memory of a self-experienced event will differ in content and quality from an account based on fabrication or imagination (Raskin & Esplin, 1991a; Steller, 1989; Undeutsch, 1989). From this notion 18 criteria were generated that are considered to be characteristics of truthful accounts. Refer to Table 1 for a list of the criteria used in the current study. For a detailed list see Raskin and Esplin (1991a).

Table 1. Means and standard deviations for each CBCA criteria as a function of event and group in Experiment 1

Group	CBCA criteria	True		Suggested		*t*	*r* effect size
		Mean	SD	Mean	SD		
Partial ($N=7$)	1. Logical structure	1.79	(.39)	.79	(.64)	2.76*	.64
	2. Production	1.07	(.79)	.57	(.67)	1.18	.28
	3. Details	1.57	(.34)	.64	(.85)	2.24*	.51
	4. Embedding	1.29	(.64)	.86	(.56)	1.28	.30
	5. Interactions	.79	(.64)	.29	(.39)	1.73	.38
	6. Reproduction	1.07	(.67)	.29	(.39)	2.75*	.52
	7. Complication	.14	(.24)	.00	(.00)	1.55	.34
	8. Unusual details	.29	(.76)	.14	(.38)	.42	.10
	9. Superfluous details	.57	(.84)	.43	(.79)	.34	.08
	11. External associations	.14	(.24)	.14	(.24)	.00	.00
	12. Subjective experience	1.20	(.91)	.79	(.86)	.76	.20
	13. Mental state	.64	(.80)	.43	(.73)	.42	.12
	14. Corrections	.21	(.39)	.29	(.57)	−.31	−.06
	15. Lack of memory	1.71	(.26)	1.93	(.19)	−1.44	−.37
	16. Raising doubts	.00	(.00)	.29	(.27)	−2.83	−.07
Full ($N=21$)	1. Logical structure	1.55	(.44)	1.48	(.51)	.51	.06
	2. Production	1.26	(.77)	1.21	(.64)	.43	.03
	3. Details	1.47	(.49)	1.24	(.58)	2.02*	.19
	4. Embedding	1.31	(.37)	1.07	(.48)	2.35*	.23
	5. Interactions	.69	(.46)	.69	(.46)	.00	.00
	6. Reproduction	.90	(.74)	1.19	(.73)	−1.61	−.17
	7. Complication	.19	(.43)	.02	(.11)	1.67	.23
	8. Unusual details	.02	(.11)	.00	(.00)	1.00	.13
	9. Superfluous details	.79	(.77)	.43	(.53)	1.70	.23
	11. External associations	.93	(.87)	.26	(.46)	3.30*	.38
	12. Subjective experience	1.55	(.76)	1.55	(.67)	.00	.00
	13. Mental state	.81	(.81)	1.09	(.68)	−1.58	−.16
	14. Corrections	.41	(.68)	.43	(.71)	−.14	−.01
	15. Lack of memory	1.62	(.59)	1.81	(.46)	−1.36	−.15
	16. Raising doubts	.12	(.27)	.33	(.48)	−2.42	−.21

Note: *Significant differences between events, $p < .05$, directional one-tailed.

DOI: 10.1002/acp

The criteria are organized into three major categories. The main category, general characteristics, is based on the premise that an account 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 (Criterion 3). The other two categories, specific contents of the statement (Criteria 4–13) and motivation-related contents (Criteria 14–18), refer to cognitive and motivational factors that are likely to be present in true accounts. For instance, a truthful person is more likely to report accurately the content of speech or conversations (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 criteria are present.

Both field and laboratory studies have tested the validity of the CBCA. In a typical study, CBCA-trained judges review accounts of events that have been determined to be true or false. These judges assess whether each of the 18 criteria are present in each account; a total score is then assigned. The main finding is that CBCA scores are higher for account of events considered true than accounts of events considered to be false (e.g. Lamb, Sternberg, Esplin, Hershkowitz, Orbach, & Hovav, 1997, Steller & Koehnken, 1989). However, a number of limitations of the CBCA have been highlighted. For example, in field studies the ground truth as to whether the allegations were true or false is not known. Thus, the results while generalizable, lack internal validity. Likewise, laboratory studies have shown several constraints in the CBCA's utility. For example, familiarity of events seems to play a more significant role in CBCA scores than veracity of events (Blandón-Gitlin, Pezdek, Rogers, & Brodie, 2005) and coaching children on the CBCA criteria can lead to increases in total scores and lower overall discriminability (Vrij, Akehurst, Soukara, & Bull, 2002).

Despite limitations of the CBCA, most studies have confirmed its underlying assumption, that more criteria are present in accounts of true events than in accounts of false events. In a qualitative review of the first 37 published CBCA studies, Vrij (2005) reported that 92% of the studies confirmed this hypothesis. Additionally, some of the criteria were more effective than others at discriminating between true and false accounts. Criterion 3 (number of details) received the most support with 80% of the studies reporting more details in true accounts than in false accounts. Unstructured production (Criterion 2), contextual embedding (Criterion 4) and reproduction of speech (Criterion 6) were found to be effective in 69% of the studies.

Vrij's (2005) review suggests that in some cases the CBCA may be useful for evaluating the veracity of accounts. Although in field studies it is unclear as to the ground truth of the cases used, it appears that in most CBCA studies the comparison has been between accounts of events that are known to be true and those which are known to be deliberately fabricated. If CBCA is to continue to be used in forensic settings, it is equally important to assess empirically whether the test can differentiate between accounts of true and suggested false events. The potential for false beliefs or even false memories is especially a concern in the context of child sexual abuse, the context in which the CBCA is primarily applied.

Some studies have provided evidence that the CBCA criteria may be useful for evaluating accounts based on illusory memories. Porter, Yuille, and Lehman (1999) asked adult participants to provide accounts of events that were (a) true, (b) fabricated and (c) suggested to have occurred. Judges analysed the accounts with a mixed set of criteria taken from different veracity assessment techniques including a few from the CBCA. These authors reported that suggested accounts showed as many *details*, but were less *coherent* than true accounts. More recently, Erdmann, Volbert, and Bohm (2004) reported

DOI: 10.1002/acp

that a few CBCA criteria were useful for discriminating between children's accounts of true and suggested events. In their study, first graders were suggested false events over multiple interviews. Analysis of children's accounts who assented to the suggestions showed that six CBCA criteria discriminated between true and suggested accounts obtained in the first interview, but only four discriminated between the accounts from the final fifth interview. Together these results suggest that the CBCA may be sensitive to differences between accounts of true and suggested events.

From a cognitive perspective, differences between accounts of true and suggested events may be expected. Johnson and Raye's (1981) reality monitoring model of memory postulates that differential encoding processes produce categories of memories' of true events that, on average, tend to be qualitatively and quantitatively different from memories of imagined events. In contrast to imagined events, true events are encoded in memory along with contextual information and within a rich elaborate network of general knowledge. Consequently, memories of true events tend to appear more real, more detailed, and more vivid than memories for imagined events. These different characteristics in the two memories representations are used as part of the process of distinguishing between true and imagined events. It would then be expected that accounts based on true memories would reveal different characteristics than accounts based on suggested memories. These differences may be captured by the CBCA criteria.

It is also possible that differences between accounts of true and suggested false events may not be revealed. Pezdek, Finger, and Hodge (1997) postulated that constructing a false memory partly involves linking the false event to a schema of similar events stored in memory. This constructive process may lead to false memories that are especially rich, compelling and seemingly real. These characteristics of the false memories, when they occur, could then be taken to indicate that they are based on real experiences. As a result, accounts of suggested events would be predicted to be qualitatively similar to accounts of true events. From this perspective, fewer differences would be expected between the CBCA scores for accounts of true and suggested false events.

Two studies were conducted to test these hypotheses and to examine the conditions under which the CBCA may or may not be effective for discriminating between accounts of true and suggested events that are believed to be true. In Experiment 1, CBCA judges analysed the transcripts of adults' descriptions of a true and a suggested childhood event. These transcripts were obtained from a study by Lindsay, Hagen, Read, Wade, and Garry (2004), which provided a unique opportunity to test the effectiveness of the CBCA in forensically relevant conditions. A second experiment was conducted so that a condition that included deliberately fabricated events could be included.

EXPERIMENT 1

Experiment 1 assessed the effectiveness of the CBCA for discriminating between accounts of true *versus* suggested false childhood events. In the Lindsay et al. (2004) study, a family-member-confederate paradigm (Loftus & Pickrell, 1995) was used to plant a false memory for a childhood event in adult participants. After providing accounts of true experiences, 45 participants were repeatedly exposed to the suggestion that they had placed a Slime toy in the desk of their first- or second-grade teacher and had been punished for doing so. The overall results showed that 62.22% (28 out of 45) of participants experienced a 'partial' (images but not memory) or 'full' memory for the target false event. Because of

DOI: 10.1002/acp

constructive processes it is likely that suggested memories are closer in content and quality to true memories, and this would be particularly so for the group categorized as having a complete or 'full' memory for the suggested event. As a result, it is predicted that if the CBCA can discriminate between true and suggested accounts, it would be more probably so for the 'partial' memory group than for the 'full' memory group. It is also predicted that specific criteria that assess memory qualities would be most sensitive to differences between the accounts regardless of memory condition. For example, number of details, coherence and contextual embedding are predicted to be present more often in true accounts than in the suggested accounts.

Method

Participants and design

Experiment 1 included the accounts of 28 participants from the Lindsay et al. (2004) study who were considered as having a 'partial' ($n = 7$) or 'full' ($n = 21$) memory for the false event. Because the purpose of this study was to assess the feasibility of the CBCA to discriminate between accounts of true events and accounts of false events believed to be true, we did not include participants who did not report a memory of the suggested event. Each participant provided accounts of a true event and a suggested false event, which made 56 transcripts available for CBCA analysis in Experiment 1. This experiment involves a 2 (true *vs*. suggested event) $\times$ 2 (partial *vs*. full memory group) mixed design, with event as the within-subject factor.

Procedure

Lindsay et al. (2004) used a family confederate paradigm to plant a memory of a false childhood event (for complete details of the procedure see Lindsay et al. pp. 2–3). The target false event was the act of putting Slime (a brightly coloured gelatinous compound) in the teacher's desk while in Grade 1 or 2. During the initial session participants were presented with narratives of two true school-related events and the false event. The first true event reportedly occurred in Grade 5 or 6 and the second in Grade 3 or 4. Half of the participants were also provided with class pictures corresponding to the grade in each event. Participants were instructed to report as much as they could remember about each of the events. A few days later participants were telephoned and encouraged to remember more about the suggested event, and a few days after that they returned to the lab for a final interview.

In the current study, the accounts of the true event experienced in Grade 3/4 were selected and compared to accounts of the Grade 1/2 false event. Two CBCA judges, blind as to the purpose of this study evaluated the 56 transcripts (28 true and 28 suggested). These judges received training at the University of British Columbia from Dr. John Yuille, a major CBCA researcher. In addition to this training, these judges have been involved in research with the CBCA and have used the instrument in forensic settings with actual legal cases in Canada. The two judges had no other role in this study except to rate the transcripts and to conduct their analyses independently of each other. The only information given to them was that some narratives involved experienced events and some involved non-experienced events. Over a period of a few weeks, the judges received batches of transcripts that included, in random order, roughly equal proportions of true and suggested accounts typed in similar format. After an initial practice with a few transcripts, judges met to discuss disagreements and further strategy. They then proceeded to independently analyze the

DOI: 10.1002/acp

transcripts. A coding sheet that included 16 CBCA criteria, a brief description of each, and a rating scale was attached to each transcript. Criteria 17 (self-depreciation) and 18 (pardoning the accused) were not used for coding because they were not applicable to the target events in this study. To be appropriate for this study, items with references to 'sexual incidents' and 'child' were reworded to indicate 'events' and 'participant'. Following the procedure from a number of CBCA studies (e.g. Boychuk, 1991; Landry & Brigham, 1992; Steller, 1989) a 0 (criterion absent) to 2 (criterion strongly present) scale was used to assess the degree to which each of the 16 criteria were present in each account.

Results and discussion

Raters' reliability

Inter-rater reliability was calculated using *Pearson Correlation, Proportion Agreement* and *Cohen's* κ (a chance-corrected measure of proportion agreement). The average Pearson r values across all criteria were $M = .57$ for true accounts and $M = .58$ for suggested accounts. Average proportion agreement was $M = .68$ for true and $M = .67$ for suggested accounts. The values for Cohen's κ were mostly in the low range across all the criteria with a few in the moderate range, resulting in an average of $M = .45$ for true accounts and $M = .29$ for suggested accounts. Overall these *item* reliabilities are in the low range and not ideal.[1] This, however, seems to be a frequent problem with CBCA studies (see Anson, Golding, & Gully, 1993; Horowitz, Lamb, Esplin, Boychuk, Krispin, & Reiter-Lavery, 1997; Sporer, 1997). It appears that in most CBCA analyses inter-rater reliability is usually moderate and particularly low with Likert scales like the one used in this study (see Sporer, 1997; Vrij, 2005). The correlations between raters' *total* CBCA scores for true and suggested accounts were large ($r = .81$ and $r = .77$, respectively) and the pattern of means in each condition was similar for both raters. Although *item* reliability was not optimal in this experiment, we nonetheless based our results on the mean of the two judges' ratings. Each judge's score was the sum of the ratings across all criteria in each transcript (maximum total score = 32). The scores from the two judges were then averaged and assigned to each transcript. Thus, the dependent variable in the following analyses is the mean CBCA score. To avoid concerns of sphericity all Fs reported in this study are the more conservative Pillai's trace statistic (see Howell, 2002, pp. 519–524 for a discussion). All r effect sizes for differences between dependent groups were calculated using procedures specified in Dunlap, Cortina, Vaslow, and Burke (1996) and Rosenthal (1991, p. 19).

Group differences on narrative length

Some researchers consider amount of information as a potential influence on CBCA analyses, with longer narratives more likely to receive higher scores than shorter narratives (e.g. Hershkowitz, Lamb, Sternberg, & Esplin, 1997). To assess if there were confounding differences in narrative length in the present study, the total number of words in the accounts were compared. A 2 (true *vs*. suggested event) × 2 (partial *vs*. full memory) mixed analysis of variance (ANOVA) conducted on the total number of words resulted in no

[1]Tables with inter-rater reliabilities for each item and inter-correlations among items are available from the first author.

DOI: 10.1002/acp

significant main effects or interaction (all F's < 1). This finding suggests that the results of the CBCA analyses in Experiment 1 cannot be attributed to narrative length.

Group differences on total CBCA scores

The main purpose of Experiment 1 was to assess whether CBCA scores differ between accounts of true and suggested false childhood events. An additional goal was to assess the conditions under which the CBCA is more or less effective for discriminating between accounts of true and suggested false childhood events. To this end, a 2 (true *vs*. suggested event) $\times$ 2 (partial *vs*. full memory) mixed ANOVA was conducted on total CBCA scores. Criterion 10 was not included in the analyses because it was never present in any transcript. The main effect of event was statistically significant, true events received higher CBCA scores ($M = 13.06$, SEM $= .87$) than suggested events ($M = 10.33$, SEM $= .79$), F (1, 26) $= 7.77$, MSE $= 10.04$, $p = .01$, $r = .77$. The main effect of memory was also significant; the full memory group received higher CBCA scores ($M = 13.21$, SEM $= .67$) than the partial memory group ($M = 10.18$, SEM $= 1.16$), F (1, 26) $= 5.10$, MSE $= 18.98$, $p = .03$, $r = .40$. Although the event by memory interaction was only marginally significant, F (1, 26) $= 3.84$, MSE $= 10.04$, $p = .06$, $\eta_p^2 = .13$, we nonetheless performed planned comparisons for the relevant groups.[2] For the partial memory group the difference in CBCA scores between true ($M = 12.50$, SEM $= 1.51$) and suggested events ($M = 7.86$, SEM $= 1.37$) was marginally significant t (6) $= 1.67$, $p = .07$, $r = .23$. For the full memory group the difference between true ($M = 13.62$, SEM $= .87$) and suggested events ($M = 12.81$, SEM $= .79$) was not statistically significant, t (20) $= 1.18$, $p = .13$, $r = .08$.

It is important to note that the effects in this experiment are relatively small, limiting the conclusions that can be drawn from these results. Nonetheless, these findings suggest that in general the CBCA is effective for discriminating between accounts of true and suggested false childhood events. However, under some conditions the effectiveness of the CBCA is significantly reduced; scores looked similar for the true and suggested accounts of participants classified as having experienced a full memory for the false childhood event.

Group differences on each of the CBCA criteria

Table 1 shows mean differences in each of the CBCA criteria as a function of event and memory group. The pattern of results shows three key aspects. First, looking at the criteria within the general characteristics (Criteria 1–3) and specific contents (Criteria 4–13) categories, it can be noted that although not statistically significant, in general, accounts of true events received higher mean scores than accounts of suggested false events. Second, the means of the suggested accounts across the majority of the criteria increased for the full memory group. These results suggest that the accounts were differentiated in the predicted direction to a greater extent for the partial memory group. Third, the mean scores of the criteria within the motivation-related (Criteria 14–16) category were higher for suggested accounts than true accounts. Because these means are not in the predicted direction, they are not interpreted as statistically significant. However, it is interesting to point out that this result should not be surprising because unlike liars, individuals with false memories are not motivated to deny weaknesses in their recollections of events.

[2]Because the key comparisons following the ANOVA were planned on theoretical grounds we chose not to correct for family wise error.

DOI: 10.1002/acp

EXPERIMENT 2

The results and interpretations from Experiment 1 should be taken with caution. Because Experiment 1 was designed to take advantage of available data, a number of experimental controls were not possible, and the generalizability of the results is thus restricted. One potential limitation of the Lindsay et al. data is that memories for the true event (Grade 3/4) were more recent than the memories for the suggested false event (Grade 1/2). Memories of recent events typically are more vivid, detailed and easily accessible than memories of distant events (Johnson, Foley, Suengas, & Raye, 1988; Sporer & Sharman, 2006). It is possible that in the constructive processes participants had more information available for their accounts of the true event than for the suggested event. This process may have worked to exaggerate differences between true and false reports and consequently the CBCA ratings. Another potential limitation is that the false event was always the Slime prank whereas the true events differed for each participant. The judges could have recognized that the Slime prank was the false event and perhaps been biased towards giving those reports lower CBCA scores. Although, this bias may have occurred, it is important to note that transcripts were randomly placed in batches that included approximately equal numbers of true and false accounts, and these were analysed by judges over a period of a few weeks. This procedure would have made it unlikely for judges to guess the true nature of the experiment. Nonetheless, the differences reported between CBCA scores for accounts of the true and suggested events could have been due in part to one or a number of confounding factors. Experiment 2 was conducted to address these concerns and to extend the generalizability of the results from Experiment 1.

In Experiment 2, recent events were manipulated to be: (a) A true experience, (b) a false experience believed to be true or (c) a false experience deliberately fabricated. This is the range of events likely to be encountered in the forensic application of the CBCA, and thus a comparison of all three is relevant. Adult participants described the three target events that were part of a sequence of activities performed in the experimental session a week earlier. In addition, participants completed the memory characteristics questionnaire (MCQ) for each of the three target events. This questionnaire, developed by Marcia Johnson and colleagues (Johnson et al., 1988) to assess qualitative and quantitative differences between memories of true and imagined events, was used in this experiment to compare participants' self-reported memory qualities with the judges' objective CBCA assessment of the accounts. Some researchers have found similarities between the CBCA approach and that of Johnson and Raye (1981) reality monitoring approach (see Sporer, 1997; Vrij, 2000).[3] Analyses with these two measures will allow an assessment of the characteristics of the two false events as compared to the true event.

Differences in CBCA scores were predicted among the three types of accounts. Specifically, it was predicted that CBCA scores would be higher for accounts of true events than for suggested and fabricated events. As with accounts of suggested events, fabricated events are predicted to lack many of the qualities of true accounts (e.g. number of details,

[3]It is important to note that in some studies the reality monitoring (RM) approach has been used as a truth detection tool (e.g. Sporer, 1997; Vrij, Akehurst, Soukara, & Bull, 2004). In these studies, judges assessed with criteria developed from the RM approach accounts of true and fabricated events to determine veracity. We did not use RM in this manner because our study's main focus is on the validity of the CBCA as a forensic tool. We obtained participants' self-reported memory qualities using the MCQ to (a) provide additional support for the judge's independent categorization of 'full' and 'partial' memories for the suggested events, (b) compare differences and similarities of memory characteristics among the three events and (c) make comparisons with the CBCA analyses.

DOI: 10.1002/acp

coherence, etc.) making them less likely to meet the CBCA criteria. Accounts of fabricated events are also expected to receive lower scores for the motivation-related criteria (14–16), which are hypothesized to be absent in deliberately manufactured lies (Vrij, 2000). No directional predictions were made for the comparisons between accounts of the suggested and fabricated events. In addition, as in Experiment 1 it is predicted that CBCA scores would be more similar for accounts of true and false events of the *full* memory group than for the *partial* memory group.

Method

Participants and design

Experiment 2 included 76 undergraduate students from a university in the Metropolitan Los Angeles area. They received class credit for their participation. The only requirement for inclusion was willingness to be videotaped during study. There were 56 women and 20 men (mean age = 19.2 years; SD = 1.29). Participants were classified based on the extent to which they experienced a memory for the suggested event (see procedure below). This classification resulted in 51 participants who were classified as having a full or partial memory for the suggested event. The design of Experiment 2 was a 2 (partial *vs*. full memory) × 3 (true *vs*. suggested *vs*. fabricated event) mixed design, with event as the within-subjects factor.

Materials and procedure

There were two sessions in this experiment. In the first session, participants performed a series of tasks and in the second session they were interviewed about some of those tasks. There were a total of 25 tasks available to be performed by participants.[4] Of those, 22 were non-target tasks and three (placing Play-Doh inside an odd-shaped box, sticking straight pins in a small scarecrow, and taking pictures of unusual animals) were the target tasks to be used as the events to be described. Most of these tasks have been successfully used in other memory suggestibility studies (e.g. Goff & Roediger, 1998; Thomas & Loftus, 2002), but are certainly much simpler than the events in Experiment 1. The target tasks were selected for unusualness to ensure that participants only described events that occurred in this study and did not simply recount episodes of familiar experiences. It is important to note that each participant actually performed only one of the three target tasks, that is, the one in the true condition; the other two target tasks were in the suggested and fabricated conditions. The three target tasks were counterbalanced across conditions to avoid confounding task with condition.

In the first session, *the performance phase*, pairs of participants in the same experimental condition were directed to a room in which two long tables were arranged with all the materials necessary to complete the tasks. Although participants performed only 23 tasks, the table included the materials to complete all 25 tasks. Each table had identical materials set up in the same order. An experimenter called aloud in random order each task to be performed. Each task was to be completed in 15 seconds; a stopwatch was used to time the duration of each activity. Participants stood in front of their assigned table and performed each task for the full 15 seconds, in some cases repeating the activity until time ran out.

[4]Contact the first author for a copy of the complete list of tasks.

DOI: 10.1002/acp

This procedure was similar to that used in other memory suggestibility studies (*cf.*, Thomas & Loftus, 2002).

One week later, *in the interview phase*, participants were interviewed individually by a different experimenter. The cover story told to participants was that the experimenter had watched him/her in the video perform a number of activities a week earlier, and selected three activities that were to be described in front of the camera on that day. To ensure participants' comfort in speaking in front of the camera, a randomly selected non-target task was introduced first for participants to describe. To establish trust with participants and ensure the believability of the suggestion, the experimenter always introduced the target events in the following order: (a) True event, (b) suggested event and (c) fabricated event. For all events, participants were told to provide as much details as possible and to try to describe each task for about 5 minutes. At the end of each description, participants were asked to complete the MCQ for each task. The MCQ includes 39 questions about memory qualities of target events that are primarily rated on a quantitative scale of 1 (little or none) to 7 (a lot) (for a copy of the MCQ, see Johnson et al., 1988).

When a participant failed to remember a target task (mostly occurred in the suggested condition), the experimenter encouraged the participant to think about contextual details such as the room where the activities took place, the table set up, and other related activities. If the participant continued to report no memory for the task, the experimenter reminded him/her that the videotape showed him/her perform the activity in the first session, and thus to try harder to retrieve the 'memory'. Another minute was given for participants to think about the task and to provide additional details that came to mind. If nothing else was remembered, the interview ended or continued with the other events.

The to-be-fabricated event was introduced last by the experimenter with the following cover story: 'There is a task that you were supposed to have performed last week but we forgot to have you do it. This task was (name of activity). Can you please pretend that you actually completed it and describe it in as much details as possible and talk for about 5 minutes'. Motivation was not manipulated in this study as is usually done in deception studies. As with the other events, after describing the fabricated event, participants completed the MCQ for this event. If participants reported no memory for the task in this condition (as expected), they were told that the MCQ probes aspects of event representation in memory, and thus they should complete it as best as their memory allowed and not to fabricate answers. Participants were then debriefed. All descriptions were transcribed verbatim.

The transcripts of the suggested event were reviewed by a trained judge (a graduate student) who assessed the extent to which participants experienced a memory for the event. The analysis involved evaluating if participants experienced (a) no memory, (b) partial memory or (c) full memories for the suggested event. To be classified as having a 'full' memory for the suggested event, participants had to report a memory as well as details for performing the target task. For a 'partial' memory participants reported details related to performing the target task. The rater judged that 51 out of the 76 participants experienced a partial ($n = 32$) or full ($n = 19$) memory for the false event. It is important to note that this classification procedure was slightly different (but comparable) to the one used by Lindsay et al. (2004). Because our tasks were much simpler and performed with other highly similar tasks, we wanted to ensure that the classification was more specifically targeted at the suggested event. The accounts of the 51 participants (153 transcripts in total) were given to the same CBCA judges from Experiment 1. Each judge evaluated the transcripts using the same procedure and scale as in Experiment 1.

DOI: 10.1002/acp

Results and discussion

Raters' reliability

Inter-rater reliability was calculated using the same three measures of agreement as in Experiment 1. The average Pearson *r* values across the criteria were, $M = .60$ for true accounts, $M = .66$ for suggested and $M = .65$ for fabricated accounts. The average proportion agreement was $M = .82$ for true accounts, $M = .81$ for suggested accounts and $M = .83$ for fabricated accounts. The average values for Cohen's κ were in the moderate range, $M = .53$, .53 and .56, respectively.[1] These *item* reliabilities are higher than in Experiment 1, especially for proportion agreement. The correlations between raters' *total* CBCA scores for true, suggested and fabricated accounts were large ($r = .76$, .68 and .70, respectively) and the pattern of means in each condition was similar for both raters. As in Experiment 1, each judge's score was the sum of the ratings across all criteria in each transcript (maximum total score = 32). The scores from the two judges were then averaged and assigned to each transcript. Thus, the dependent variable in the following analyses is the mean CBCA score.

Group differences in narrative length

A 2 (partial *vs*. full memory) × 3 (true *vs*. suggested *vs*. fabricated event) mixed ANOVA conducted on the total number of words resulted in a significant main effect of event, $F(2, 48) = 7.21$, MSE = 8052.78, $p = .01$, $\eta_p^2 = .23$. This result was due to accounts of fabricated events ($M = 362.96$, SEM = 24.01) containing significantly fewer words than accounts of true events ($M = 404.67$, SEM = 21.47, $t(50) = 1.83$, $p < .05$, $r = .11$) and suggested events ($M = 428.24$, SEM = 19.16, $t(50) = 3.55$, $p < .01$, $r = .19$). The main effect of memory type ($F(1, 49) = 2.33$, MSE = 54 003.19, $p > .05$, $\eta_p^2 = .05$) and interaction ($F(2, 48) = 1.63$, MSE = 8052.78, $p > .05$, $\eta_p^2 = .06$) were not significant. Although accounts of fabricated events contained fewer words than the other two types of accounts, there was no significant difference between accounts of true and suggested events, the most critical comparison in this study.

Group differences on total CBCA scores

The main purpose of Experiment 2 was to assess the effectiveness of the CBCA for discriminating among accounts of true, suggested and deliberately fabricated events. The second purpose was to assess conditions under which CBCA is more or less effective. Again, criterion 10 was excluded from analyses because it was not present in the transcripts. A 2 (partial *vs*. full memory) × 3 (true *vs*. suggested *vs*. fabricated event) mixed ANOVA showed a main effect of event, $F(2, 48) = 5.94$, MSE = 13.37, $p < .01$, $\eta_p^2 = .20$. Further analyses showed (a) accounts of true events received significantly higher CBCA scores ($M = 11.39$, SEM = .37) than suggested events ($M = 9.58$, SEM = 44); $t(50) = 3.81$, $p < .05$, $r = .30$, (b) CBCA scores were slightly higher for true events ($M = 11.39$, SEM = .37) than fabricated events ($M = 10.62$, SEM = .40) and this difference was marginally significant, $t(50) = 1.64$, $p = .05$, $r = .18$ and (c) the difference between accounts of suggested and fabricated events was also statistically significant, $t(50) = 1.79$, $p < .05$, $r = .25$. The main effect of memory ($F(1, 48) = 1.78$, MSE = 11.54, $p > .05$, $\eta_p^2 = .04$) and the interaction ($F(2, 48) = 1.98$, MSE = 6.68, $p > .05$, $\eta_p^2 = .08$) were not significant.

Although, the interaction was not statistically significant, planned comparisons were nonetheless conducted between target groups, motivated by the pattern of findings in Experiment 1. For the *partial* memory group the differences in CBCA scores between true ($M = 12.02$, SEM = .41) and suggested events ($M = 9.59$, SEM = .53) was significant *t*

DOI: 10.1002/acp

(31) = 3.95, $p < .01$, $r = .42$. Similarly, CBCA scores were higher for true events ($M = 12.02$, SEM = .41) than fabricated events ($M = 10.88$, SEM = .43), t (31) = 2.14, $p < .05$, $r = .24$. The difference in CBCA scores between accounts of suggested and fabricated events was also statistically significant, t (31) = 1.81, $p < .05$, $r = .23$. For the *full* memory group none of the planned pairwise comparisons were significant (largest $t = 1.12$, $p > .05$).

Group differences on each of the CBCA criteria

Table 2 shows mean differences in each of the CBCA criteria as a function of event and memory Group. Three aspects of these data can be noted. First, consistent with the results from Experiment 1, the partial memory group's true and suggested accounts could be differentiated on the basis of some of the CBCA criteria. Second, the results show that the majority of means for the full memory group increased, suggesting a reduction in CBCA's

Table 2. Means and standard deviations for each CBCA criteria as a function of event and group in Experiment 2

Group	CBCA criteria	True		Suggested		Fabricated		F	η_p^2
		Mean	SD	Mean	SD	Mean	SD		
Partial ($N = 32$)	1. Logical structure	1.77[a]	(.38)	.73[ab]	(.71)	1.72[b]	(.44)	29.68*	.66
	2. Production	1.75[a]	(.36)	.80[ab]	(.75)	1.67[b]	(.49)	25.93*	.63
	3. Details	1.59[a]	(.56)	1.03[ab]	(.61)	1.51[b]	(.62)	11.10*	.43
	4. Embedding	.92	(.61))	.87[a]	(.55)	1.06[a]	(.44)	1.90*	.11
	5. Interactions	.33[ab]	(.41)	.19[a]	(.31)	.19[b]	(.31)	1.62	.10
	6. Reproduction	.64[a]	(.60)	.37[ab]	(.42)	.64[b]	(.69)	4.89*	.25
	7. Complication	.66	(.75)	.53	(.67)	.56	(.64)	.21	.01
	8. Unusual details	0.00	(.00)	.03	(.18)	.05	(.20)	1.44	.09
	9. Superfluous details	.42	(.57)	.83	(.55)	.41	(.48)	6.92*	.32
	11. External associations	.66	(.82)	.39	(.67)	.38	(.68)	2.34	.14
	12. Subjective experience	1.73[ab]	(.60)	1.30[a]	(.89)	1.05[b]	(.91)	5.31*	.26
	13. Mental state	.14	(.43)	.14	(.43)	.17	(.55)	.06	.00
	14. Corrections	.14	(.36)	.28	(.60)	.33	(.68)	1.28	.08
	15. Lack of memory	1.17	(.89)	1.72[a]	(.63)	.86[a]	(.83)	13.87*	.48
	16. Raising doubts	.03	(.12)	.06	(.21)	0.00	(.00)	1.55	.09
Full ($N = 19$)	1. Logical structure	1.55[a]	(.58)	.89[ab]	(.81)	1.66[b]	(.47)	7.16*	.46
	2. Production	1.50[a]	(.55)	.95[ab]	(.85)	1.45[b]	(.47)	4.35*	.34
	3. Details	1.47[a]	(.66)	1.18[a]	(.56)	1.37	(.57)	2.10	.20
	4. Embedding	.76	(.54)	.95	(.47)	1.16[a]	(.53)	6.19*	.42
	5. Interactions	.18[ab]	(.25)	.53[ab]	(.54)	.24[b]	(.31)	3.96*	.32
	6. Reproduction	.68[a]	(.51)	.47[b]	(.49)	.18[ab]	(.45)	7.35*	.46
	7. Complication	.63	(.81)	.63	(.78)	.37	(.64)	.76	.08
	8. Unusual details	0.00	(.00)	.00	(.00)	.13	(.28)	4.17	.19
	9. Superfluous details	.47	(.57)	.66	(.57)	.47	(.72)	1.17	.12
	11. External associations	.18	(.42)	.24	(.54)	.58[a]	(.85)	2.08	.20
	12. Subjective experience	1.18[ab]	(.80)	.74[a]	(.90)	1.13	(.88)	1.65	.16
	13. Mental state	0.00	(.00)	.18	(.48)	.26	(.65)	2.34	.28
	14. Corrections	.32	(.67)	.37	(.68)	.37	(.60)	.05	.01
	15. Lack of memory	1.24	(.95)	1.55[a]	(.64)	.95[a]	(1.03)	2.26	.21
	16. Raising doubts	.03	(.12)	.08	(.25)	.00	(.00)	1.47	.15

Note: *Significant differences among events, $p < .05$; [a/b]means sharing the same superscript are statistically different, $p < .05$, directional one-tailed.

DOI: 10.1002/acp

discriminative power in this condition. Third, for the most part, fabricated events received similar CBCA scores to the true events. One explanation for this finding is that because participants described the fabricated event last in the interview sequence, it is possible that they were more practiced and thus more likely to produce the type of information that elicits higher CBCA criteria. Similar findings were reported by Sporer (1997), when participants were given extra time to prepare a false story; CBCA scores of invented accounts were not significantly different from accounts of true events.

Group differences on the MCQ self-ratings

Participants' self-reported MCQ ratings were next analysed. Sporer (1997, 2004) factor analysed the 39 questions of the MCQ effectively creating eight criteria. The responses to the 39 MCQ questions in this study were clustered into these eight criteria. Table 3 presents these results. For 5 out of the 8 criteria a 2 (partial *vs*. full memory) × 3 (true *vs*. suggested *vs*. fabricated event) revealed a significant interaction. These were for c*larity*, F (2, 47) = 12.29, MSE = 1.86, $p < .01$, $\eta_p^2 = .34$; s*ensory experience* F (2, 47) = 4.97, MSE = .61, $p < .01$, $\eta_p^2 = .18$; *emotions*, F (2, 44) = 4.57, MSE = .49, $p < .05$, $\eta_p^2 = .17$; *story reconstructability*, F (2, 47) = 3.14, MSE = .49, $p < .05$, $\eta_p^2 = .12$; *cognitive operations* F (2, 47) = 8.12, MSE = .44, $p < .01$, $\eta_p^2 = .26$. The *spatial information*, *time information* and *realism* criteria did not show a significant interaction, all $Fs < 2.00$.

Planned simple effects analyses showed that for the partial memory group, 7 out of 8 one-way within subjects ANOVAS were significant. These results are presented in the upper panel of Table 3. Furthermore, planned comparisons revealed that for the most part, true events were rated significantly higher than suggested events. Ratings of fabricated

Table 3. Mean self-ratings on the MCQ criteria for each event as a function of group in Experiment 2 (scale 1–7; higher numbers indicate more of the stated criteria)

Group	MCQ criteria	True		Suggested		Fabricated		F	η_p^2
		Mean	SD	Mean	SD	Mean	SD		
Partial ($N = 32$)	1. Clarity	4.41^{a}	(1.30)	219^{a}	(1.34)	3.20^{a}	(1.89)	18.28^{**}	.56
	2. Sensory experience	3.15^{a}	(.73)	2.28^{ab}	(.70)	2.74^{b}	(1.36)	9.78^{**}	.40
	3. Spatial information	5.59^{a}	(1.02)	5.17^{ab}	(1.39)	5.13^{b}	(1.77)	3.99^{*}	.21
	4. Time information	5.99	(1.18)	5.96	(1.14)	5.75	(1.68)	.61	.04
	5. Emotions	3.86^{ab}	(1.02)	3.06^{a}	(.74)	3.10^{b}	(1.03)	9.16^{**}	.40
	6. Story reconstructability	2.48^{a}	(.88)	1.90^{ab}	(.99)	2.23^{b}	(1.06)	6.80^{**}	.32
	7. Realism	3.88^{a}	(2.01)	3.34	(1.99)	2.84^{a}	(2.25)	3.67^{*}	.20
	8. Cognitive operations	2.83	(.92)	2.07	(.86)	2.16	(1.04)	12.20^{**}	.46
Full ($N = 19$)	1. Clarity	5.21^{a}	(1.05)	4.24^{a}	(1.51)	2.72^{a}	(1.54)	19.96^{**}	.70
	2. Sensory experience	2.89	(.81)	2.81	(0.84)	2.61	(1.18)	.72	.08
	3. Spatial information	5.97^{a}	(0.75)	5.89	(0.88)	5.18^{a}	(1.97)	1.78	.17
	4. Time information	6.29	(0.98)	6.15	(1.27)	5.54	(2.00)	1.91	.18
	5. Emotions	3.66^{a}	(0.97)	3.48^{b}	(0.97)	2.71^{ab}	(1.07)	6.99^{*}	.47
	6. Story reconstructability	2.75^{ab}	(1.00)	2.43^{a}	(0.89)	2.21^{b}	(1.09)	2.18	.20
	7. Realism	4.42^{a}	(2.10)	4.32^{b}	(2.14)	3.05^{ab}	(2.17)	4.77	.36
	8. Cognitive operations	2.95	(1.03)	2.74	(0.94)	2.04	(1.02)	5.50^{*}	.39

Note: *Significant differences among events, $p < .05$; **significant differences among events $p < .01$; $^{a/b}$means sharing the same superscript are statistically different, $p < .05$, directional one-tailed.

DOI: 10.1002/acp

events were in general lower than that of true events, but not always significantly so. These results suggest that self-ratings on the MCQ can effectively differentiate between memories of experienced and non-experienced events.

The same planned analyses were performed on the full memory group. As can be seen in the lower panel of Table 3, the most salient result is that the MCQ ratings for the suggested event are higher across all the criteria as compared to the ratings of the partial memory group. This increase in ratings effectively reduced the number of comparisons that reached significance for the full memory group. Of the eight one-way within subjects ANOVAS performed, four were not significant. Additionally, whereas five of the planned pair-wise comparisons between ratings of the *true* and *suggested* events were significant in the partial group, only two of the same comparisons were statistically significant (MCQ Criteria 1 and 6) for the full memory group.

The findings from the MCQ data support the results of the CBCA analyses. Memories for the true events received higher MCQ ratings than memories for the suggested events of the partial memory group. The true event memories of this latter group were in general perceived as clearer, easily reconstructed, with more emotion, sensory and spatial details than memories for the suggested event. This difference was substantially reduced for the full memory group. Together these results suggest that the CBCA and the MCQ may tap similar constructs even though one is an objective content assessment of accounts and the other is a subjective self-assessment of memories for events.

It is important to note that Criterion 8 (cognitive operations) was expected to be present more often in the ratings of false events than true events. This is because the reality monitoring framework postulates that compared to true memories, imagined memories involve more information about the cognitive process that produced them (Johnson et al., 1988). Our results showed that MCQ ratings were higher for the true events than for the suggested and fabricated events. This suggests that the cognitive operation criterion as applied here may not optimally capture the predicted differences between accounts (see Sporer & Sharman, 2006, p. 844).

GENERAL DISCUSSION

There are two key findings in this study. First, the results from both experiments support the Undeutsch hypothesis, which postulates average content differences between accounts of events that are based on true experiences and account of events that are based on fantasy or invention (Steller, 1989). In Experiment 1, accounts of true childhood events received higher CBCA scores than accounts of suggested childhood events, and in Experiment 2 this result was replicated with recent events. The results also revealed that the effect is mainly due to the CBCA criteria that tap general characteristics of events and to some extent to other criteria tapping specific characteristics of events. This latter result suggests important differences in the sensitivity of the various CBCA criteria.

Why would accounts of experienced events be expected to have characteristics that differ from accounts of non-experienced events? According to Johnson et al. (1988), memories for experienced events are stored and embedded in memory within an elaborate network of information that typically includes a large quantity of perceptual details (e.g. colour, sound and smell) and contextual information (e.g. time and place). On the other hand, memories for imagined and otherwise non-experienced events typically include less perceptual and contextual information and instead have more information about the

DOI: 10.1002/acp

cognitive process that produced them. Based on this notion it was hypothesized that accounts based on true and false memories would show a predictable pattern of difference. These differences appear to be captured by some of the CBCA criteria as well as the MCQ criteria. Across the two experiments in this study, accounts of true events were rated by CBCA judges to be more logical, more coherent and to have more details than accounts of suggested events. Additionally, in Experiment 2, participants rated their memories for the true events to be clearer, to contain more emotion, sensory and spatial information and more easily reconstructed than their memories for the suggest event. Although there is certainly overlap between these events (e.g. some true events were rated as less vivid than suggested events) the means indicate a general sensitivity of CBCA and MCQ criteria for discriminating between true and false events.

The second key finding from this study is that the small discriminative power of the CBCA was further reduced in the accounts of the full memory group. The CBCA scores of the true and suggested events were similar for this group. Why did this occur? Pezdek et al. (1997) and Pezdek and Hinz (2002) and others (e.g. Loftus & Bernstein, 2004) concluded that false memories are partly constructed by combining information from related true events with details of the suggested event. It is possible that this process occurred to a greater extent in the full memory group with the resulting accounts containing more perceptual details that met more of the CBCA criteria.

It is important to note that many of the CBCA criteria within the specific contents category of the CBCA had poor discriminative power. Perhaps this was because of the age of the memories for events in Experiment 1 and because of the nature of the events in Experiment 2. Because of normal forgetting, accounts based on memories for distant events, would be expected to include fewer details than accounts of recent events. The fewer key details in the accounts, the lower the probability that CBCA criteria would be met. Johnson et al. (1988) reported significant differences between the memory characteristics of recent true and imagined events, but not between true and imagined distant (childhood) events (see Sporer & Sharman, 2006 for similar results). Furthermore, in Experiment 2, the lower discriminative power of the specific contents category could have been due to the fact that the tasks that served as the target events were interrelated, sharing many of the same particular features measured by the criteria in this category. If something unusual occurred during the experimental session, the probability that participants would report this detail was the same for all three events.

Although some researchers have begun to differentially weight the presence of particular CBCA criteria in statements (e.g. Erdmann et al., 2004), at the moment there are no formal structured procedures for weighing the criteria when performing CBCA evaluations. The results of this study, together with the findings of comprehensive CBCA reviews (e.g. Vrij, 2005), suggest that this is a limitation in the current form of the CBCA. In both experiments in this study, the criteria within the general characteristics category were more effective at discriminating between accounts of true and suggested events. These criteria take into consideration holistic aspects of the account that are considered especially important by some researchers (Pezdek et al., 2004). A promising direction for future CBCA research is developing a formal weighting system to increase the accuracy and enhance sensitivity of the evaluations.

The goal of this study was to assess the feasibility of the CBCA for discriminating between accounts of true events and false events believed to be true. This is a situation likely to be encountered in the forensic application of this technique, yet up to now it has not been systematically examined. Although particular CBCA criteria may be potentially

DOI: 10.1002/acp

sensitive to differences between accounts of true and suggested events, the research findings reported here shows that the utility of the CBCA is severely constrained as a forensic tool. For subsets of participants in this study—those who developed specially compelling false memories—the CBCA did not discriminate between accounts of the true and suggested false events. In other words, for individuals who believe they experienced a suggested event, their accounts are more likely to appear real, and thus classified as being true by CBCA standards. Although, this is not a failure of the CBCA per se because its main purpose is to discriminate between truth tellers and liars, this study shows that it may be potentially biased when the deception is subjectively true.

ACKNOWLEDGEMENTS

This research was partially supported by grants from the John Randolph Haynes Foundation to Iris Blandon-Gitlin and from the National Science Foundation to Kathy Pezdek, 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 Haynes Foundation or the National Science Foundation.

REFERENCES

Anson, D. A., Golding, S. L., & Gully, K. J. (1993). Child sexual abuse allegations: Reliability of criterion-based content analysis. *Law and Human Behavior*, *17*, 331–341.

Blandón-Gitlin, I., Pezdek, K., Rogers, R., & Brodie, L. (2005). The effects of event familiarity on criterion-based content analysis ratings: An experimental study. *Law and Human Behavior*, *29*, 187–197.

Boychuk, 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.

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods*, *1*, 170–177.

Erdmann, K., Volbert, R., & Bohm, C. (2004). Children report suggested events even when interviewed in a non-suggestive manner: What are its implications for credibility assessment? *Applied Cognitive Psychology*, *18*, 561–589.

Goff, L. M., & Roediger, H. L., III (1998). Imagination inflation for action events: Repeated imaginings lead to illusory recollections. *Memory and Cognition*, *26*, 20–33.

Hershkowitz, I., Lamb, M. E., Sternberg, K. J., & Esplin, P. W. (1997). The relationships among interviewer utterance type, CBCA scores and the richness of children's responses. *Legal and Criminological Psychology*, *2*, 169–176.

Horowitz, S. W., Lamb, M. E., Esplin, P. W., Boychuk, T. D., Krispin, O., & Reiter-Lavery, L. (1997). Reliability of criteria-based content analysis of child witness statements. *Legal and Criminological Psychology*, *2*, 11–21.

Howell, D. C. (2002). *Statistical methods for psychology* (5th ed.). Boston: Thomson Learning.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, *88*, 67–85.

Johnson, M. K., Foley, M. A., Suengas, A. G., & Raye, C. L. (1988). Phenomenal characteristics of memories for perceived and imagined autobiographical events. *Journal of Experimental Psychology: General*, *177*, 371–376.

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). Criterion-based content analysis: A field validation study. *Child Abuse and Neglect*, *21*, 255–264.

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.

DOI: 10.1002/acp

Lindsay, D. S., Hagen, L., Read, J. D., Wade, K. A., & Garry, M. (2004). True photographs and false memories. *Psychological Science*, *15*, 149–154.

Loftus, E. F., & Bernstein, D. M. (2005). Rich false memories: The royal road to success. In A. J. F. Healy (Ed.), *Experimental cognitive psychology and its implications*. *Decade of Behavior* (pp. 101–113). Washington DC, US: American Psychological Association.

Loftus, E. F., & Pickrell, J. (1995). The formation of false memories. *Psychiatric Annals*, *25*, 720–725.

Pezdek, K., & Hinz, T. (2002). The construction of false events in memory. In H. L. Westcott, G. M. Davis, & R. H. C. Bull (Eds.), *Children's testimony: A handbook of psychological research and forensic practice*. New York: John Wiley & Sons.

Pezdek, K., Finger, K., & Hodge, D. (1997). Planting false childhood memories: The role of event plausibility. *Psychological Science*, *8*, 437–441.

Pezdek, K., Morrow, A., Blandon-Gitlin, I., Goodman, G. S., Quas, J. A., Saywitz, K. J., et al. (2004). Detecting deception in children: The effect of event familiarity on Criterion-Based Content Analysis ratings. *Journal of Applied Psychology*, *89*, 119–126.

Porter, S., Yuille, J. C., & Lehman, D. R. (1999). The nature of real, implanted, and Fabricated memories for emotional childhood events: Implications for the recovered memory debate. *Law and Human Behavior*, *2*, 517–537.

Raskin, D. C., & Esplin, P. W. (1991a). Statement validity assessment: Interview procedures and content analysis of children's statements of sexual abuse. *Behavioral Assessment*, *12*, 265–291.

Raskin, D. C., & Esplin, P. W. (1991b). 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.

Rosenthal, R. (1991). *Meta-analytic procedures for social research*. CA, US: Sage Publications.

Sporer, S. L. (1997). The less travelled road to truth: Verbal cues in deception in accounts of fabricated and self-experienced events. *Applied Cognitive Psychology*, *11*, 373–397.

Sporer, S. L. (2004). Reality monitoring and detection of deception. In P. A. Granhag, & L. A. Stromwall (Eds.), *The detection of deception in forensic contexts* (pp. 64–102). Cambridge, UK: Cambridge University Press.

Sporer, S. L., & Sharman, S. J. (2006). Should I believe this? Reality monitoring of accounts of self-experienced and invented recent and distant autobiographical events. *Applied Cognitive Psychology*, *20*, 837–854.

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., & Koehnken, 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.

Thomas, A. K., & Loftus, E. F. (2002). Creating bizarre false memories through imagination. *Memory and Cognition*, *30*, 423–431.

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.

Undeutsch, U. (1989). The development of statement reality analysis. In J. C. Yuille (Ed.), *Credibility assessment* (pp. 101–121). Dordrecht, Netherlands: Kluwer Academic Publishers.

Vrij, A. (2000). *Detecting lies and deceit: The psychology of lying and the implications for professional practice*. New York: John Wiley & Sons.

Vrij, A. (2005). Criteria-based content analysis: A qualitative review of the first 37 studies. *Psychology, Public Policy and Law*, *11*, 3–41.

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., Akehurst, L., Soukara, S., & Bull, R. (2004). Let me inform you how to tell a convincing story: CBCA and reality monitoring scores as a function of age, coaching, and deception. *Canadian Journal of Behavioral Science*, *36*, 113–126.

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

DOI: 10.1002/acp